DOCUMENT RESUME

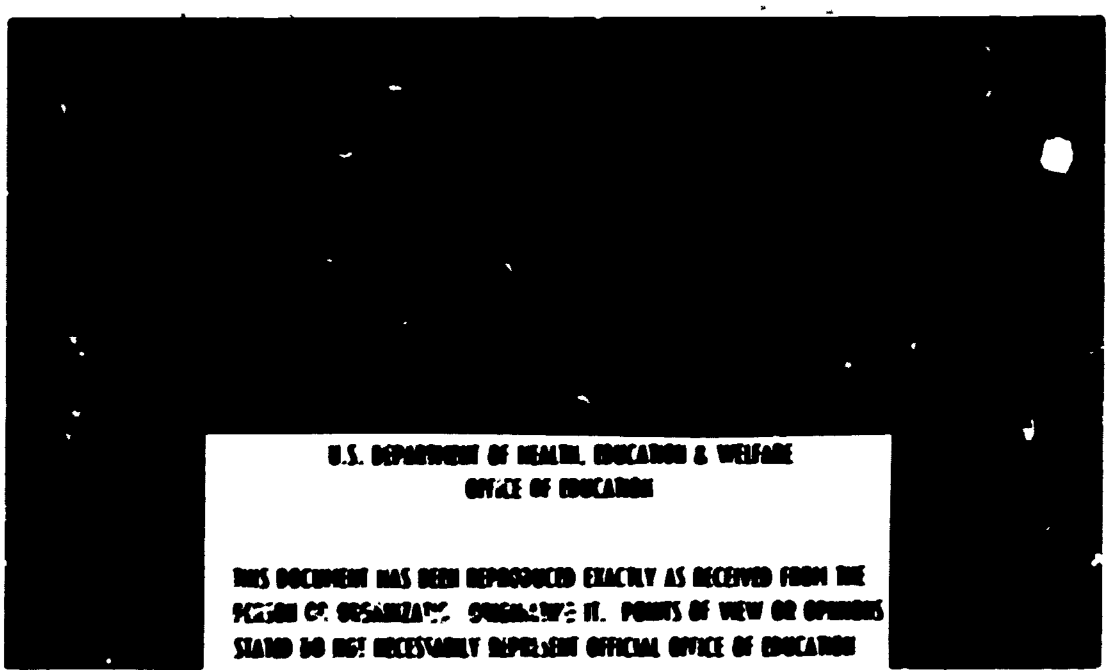
ED 038 293	24	SE 008 177

AUTHOR	Meyer, Rochelle Wilson
TITLE	The Identification and Encouragement of Mathematical Creativity in First Grade Students, (Part 2) (Chapter 5 to Conclusion).
INSTITUTION	Wisconsin Univ., Madison. Research and Development Center for Cognitive Learning.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C. Bureau of Research.
REPORT NO	TR-112
BUREAU NO	BR-5-0216
PUB DATE	Jan 70
CONTRACT	OEC-5-10-154
NOTE	96p.
EDRS PRICE	EDRS Price MF-$0.50 HC-$4.90
DESCRIPTORS	*Creative Thinking, Creativity, Doctoral Theses, *Elementary School Mathematics, *Grade 1, *Learning, Mathematical Concepts, *Research

ABSTRACT

Reported is research involving the development and testing or a program designed to encourage individual creative mathematical activity in first grade students. Initially, some characteristics of the creative process and creative thinking were examined and six criteria describing certain aspects of mathematical creativity were identified and validated. An instrument used to measure observable mathematical creativity was designed in order to test the program. An experiment was conducted to determine the effects of participation in this program on mathematical creativity. Two hypotheses were formulated H1: "Participation in the program will increase a student's observable mathematical creativity"; H2: "Participation in the program will not affect a student's performance on a test of general creative ability." Part II includes Chapters V-VII of this report. Chapter V describes the test instrument which was developed. Chapter VI discusses the design of the experiment and the statistical analyses made on the data. Conclusions , and implications for future study are found in Chapter VII. The experimental program offers evidence that under certain suitable conditions first grade students can exhibit behavior which satisfies all the criteria describing aspects of mathematically creative ability. (FL)

BR 5-0216
PT II
SE



No. 112 (Part II) (Chapter V to Conclusion)

THE IDENTIFICATION AND ENCOURAGEMENT OF MATHEMATICAL CREATIVITY IN FIRST GRADE STUDENTS

Report from the Project on
Analysis of Mathematics Instruction

U.S. Office of Education Center No. C-03 Contract OE 5-10-154

BR-5-0216
TR-112 (Part II)
PA-24
OE/BR

Technical Report No. 112 (Part II) (Chapter V to Conclusion)

THE IDENTIFICATION AND ENCOURAGEMENT OF MATHEMATICAL CREATIVITY IN FIRST GRADE STUDENTS

Report from the Project on Analysis of Mathematics Instruction

By Rochelle Wilson Meyer

John G. Harvey, Professor of Curriculum & Instruction & Mathematics
Chairman of the Examing Committee

John G. Harvey and Thomas A. Romberg, Principal Investigators

Wisconsin Research and Development
Center for Cognitive Learning
The University of Wisconsin
Madison, Wisconsin

January 1970

This Technical Report is a doctoral dissertation reporting research supported by the Wisconsin Research and Development Center for Cognitive Learning. Since it has been approved by a University Examining Committee, it has not been reviewed by the Center. It is published by the Center as a record of some of the Center's activities and as a service to the student. The bound original is in The University of Wisconsin Memorial Library.

Published by the Wisconsin Research and Development Center for Cognitive Learning, supported in part as a research and development center by funds from the United States Office of Education, Department of Health, Education, and Welfare. The opinions expressed herein do not necessarily reflect the position or policy of the Office of Education and no official endorsement by the Office of Education should be inferred.

Center No. C 03 / Contract OE 5-10-154

NATIONAL EVALUATION COMMITTEE

Samuel Brownell
Professor of Urban Education
Graduate School
Yale University

Henry Chauncey
President
Educational Testing Service

Elizabeth Koontz
President
National Education Association

Patrick Suppes
Professor
Department of Mathematics
Stanford University

Launor F. Carter
Senior Vice President on
Technology and Development
System Development Corporation

Martin Deutsch
Director, Institute for
Developmental Studies
New York Medical College

Roderick McPhee
President
Punahou School, Honolulu

***Benton J. Underwood**
Professor
Department of Psychology
Northwestern University

Francis S. Chase
Professor
Department of Education
University of Chicago

Jack Edling
Director, Teaching Research
Division
Oregon State System of Higher
Education

G. Wesley Sowards
Director, Elementary Education
Florida State University

UNIVERSITY POLICY REVIEW BOARD

Leonard Berkowitz
Chairman
Department of Psychology

John Guy Fowlkes
Director
Wisconsin Improvement Program

Herbert J. Klausmeier
Director, R & D Center
Professor of Educational
Psychology

M. Crawford Young
Associate Dean
The Graduate School

Archie A. Buchmiller
Deputy State Superintendent
Department of Public Instruction

Robert E. Grinder
Chairman
Department of Educational
Psychology

Donald J. McCarty
Dean
School of Education

***James W. Cleary**
Vice Chancellor for Academic
Affairs

H. Clifton Hutchins
Chairman
Department of Curriculum and
Instruction

Ira Sharkansky
Associate Professor of Political
Science

Leon D. Epstein
Dean
College of Letters and Science

Clauston Jenkins
Assistant Director
Coordinating Committee for
Higher Education

Henry C. Weinlick
Executive Secretary
Wisconsin Education Association

EXECUTIVE COMMITTEE

Edgar F. Borgatta
Brittingham Professor of
Sociology

Russell J. Hosler
Professor of Curriculum and
Instruction and of Business

Wayne Otto
Professor of Curriculum and
Instruction (Reading)

Richard L. Venezky
Assistant Professor of English
and of Computer Sciences

Max R. Goodson
Professor of Educational Policy
Studies

***Herbert J. Klausmeier**
Director, R & D Center
Professor of Educational
Psychology

Robert G. Petzold
Associate Dean of the School
of Education
Professor of Curriculum and
Instruction and of Music

FACULTY OF PRINCIPAL INVESTIGATORS

Ronald R. Allen
Associate Professor of Speech
and of Curriculum and
Instruction

Gary A. Davis
Associate Professor of
Educational Psychology

Max R. Goodson
Professor of Educational Policy
Studies

Richard G. Morrow
Assistant Professor of
Educational Administration

Vernon L. Allen
Associate Professor of Psychology
(On leave 1968-69)

M. Vere DeVault
Professor of Curriculum and
Instruction (Mathematics)

Warren O. Hagstrom
Professor of Sociology

Wayne Otto
Professor of Curriculum and
Instruction (Reading)

Nathan S. Blount
Associate Professor of English
and of Curriculum and
Instruction

Frank H. Farley
Assistant Professor of
Educational Psychology

John G. Harvey
Associate Professor of
Mathematics and Curriculum
and Instruction

Milton O. Pella
Professor of Curriculum and
Instruction (Science)

Robert C. Calfee
Associate Professor of Psychology

John Guy Fowlkes (Advisor)
Professor of Educational
Administration
Director of the Wisconsin
Improvement Program

Herbert J. Klausmeier
Director, R & D Center
Professor of Educational
Psychology

Thomas A. Romberg
Assistant Professor of
Mathematics and of
Curriculum and Instruction

Robert E. Davidson
Assistant Professor of
Educational Psychology

Lester S. Golub
Lecturer in Curriculum and
Instruction and in English

Burton W. Kreitlow
Professor of Educational Policy
Studies and of Agricultural
and Extension Education

Richard L. Venezky
Assistant Professor of English
and of Computer Sciences

MANAGEMENT COUNCIL

***Herbert J. Klausmeier**
Director, R & D Center
Acting Director, Program 1

Thomas A. Romberg
Director
Programs 2 and 3

James E. Walter
Director
Dissemination Section

Dan G. Woolpert
Director
Operations and Business

Mary R. Quilling
Director
Technical Section

* COMMITTEE CHAIRMAN

STATEMENT OF FOCUS

The Wisconsin Research and Development Center for Cognitive Learning focuses on contributing to a better understanding of cognitive learning by children and youth and to the improvement of related educational practices. The strategy for research and development is comprehensive. It includes basic research to generate new knowledge about the conditions and processes of learning and about the processes of instruction, and the subsequent development of research-based instructional materials, many of which are designed for use by teachers and others for use by students. These materials are tested and refined in school settings. Throughout these operations behavioral scientists, curriculum experts, academic scholars, and school people interact, insuring that the results of Center activities are based soundly on knowledge of subject matter and cognitive learning and that they are applied to the improvement of educational practice.

This Technical Report is from Phase 2 of the Project on Prototypic Instructional Systems in Elementary Mathematics in Program 2. General objectives of the Program are to establish rationale and strategy for developing instructional systems, to identify sequences of concepts and cognitive skills, to develop assessment procedures for those concepts and skills, to identify or develop instructional materials associated with the concepts and cognitive skills, and to generate new knowledge about instructional procedures. Contributing to the Program objectives, the Mathematics Project, Phase 1, is developing and testing a televised course in arithmetic for Grades 1-6 which provides not only a complete program of instruction for the pupils but also inservice training for teachers. Phase 2 has a long-term goal of providing an individually guided instructional program in elementary mathematics. Preliminary activities include identifying instructional objectives, student activities, teacher activities materials, and assessment procedures for integration into a total mathematics curriculum. The third phase focuses on the development of a computer system for managing individually guided instruction in mathematics and on a later extension of the system's applicability.

To my parents

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

LIST OF TABLES

LIST OF TABLES (CONTINUED)

ACKNOWLEDGEMENTS

Many people have helped in the research reported in this thesis. Although it is impossible to mention everyone here, some people deserve special recognition for their assistance.

Dr. John G. Harvey has been what good major professors are supposed to be--he has been of valuable help to me both in formulating the research and in writing the thesis. I was not once but twice fortunate; in addition to Dr. Harvey I was able to benefit from the advice of Dr. Thomas A. Romberg, who has given of his time and ideas as if I were his own student. These two, plus Dr. Joshua Chover and Dr. Gary A. Davis, read the drafts of the thesis and made many good suggestions.

Two pilot studies were graciously accepted into the Prospect Street Elementary School, Lake Mills, Wisconsin. For their cooperation I thank Mr. Ronald Hering, Director of Research Activities; Mr. Thomas Block, Principal; Mrs. Linda Bender, Mrs. Marianna Buchanan, Miss Cheryl Hagan, and Mrs. Inez Shultz, teachers. A second gracious acceptance allowed the experiment to be conducted in the Poynette Elementary School, Poynette, Wisconsin. I am indebted to Mr. Gerald Makie, Superintendent; Mr. Glenn Porterfiled, Principal; Mr. Charles Tucker, Elementary Supervisor; and Mrs. Hansen, the teacher, for their warm welcome and cooperation. Mr. George Glasrud located these two friendly school districts and made the initial contacts.

For taking time out of their busy schedules to participate in the face validation I thank Dr. Richard A. Brualdi, Dr. Joshua Chover, Dr. Donald Crowe, Dr. Simon Hellerstein, Dr. Mary Ellen Rudin, Dr. Hans Schneider, and Dr. Melvin C. Thornton.

Mr. John McFee did the videotaping and editing and managed to remain cheerful and interesting through it all.

Mrs. Carolyn Gornowicz did such a fine job of teaching during the experiment that I could relax and enjoy the students, knowing that they were in good hands.

Mr. Thomas Fischbach helped in some difficult problems of statistical analysis.

Mrs. Cindy Gaa not only typed the manuscript but pointed out missing verbs and commas along the way.

This research was supported by the United States Office of Education, Department of Health, Education, and Welfare, under the provisions of the Cooperative Research Program, through a contract with the Wisconsin Research and Development Center for Cognitive Learning (Center No. C-03/Contract OE 5-10-154).

My husband, Walter Meyer, has helped in ways that words cannot adequately express--ways that are in some sense so subtle that I find new aspects creeping into my awareness every day. All this while he too was writing a thesis.

And last, the students who took part in the pilot studies and the experiment and whose honest and creative reactions to the mathematics problems were wonderful and rewarding to watch.

ABSTRACT

The importance of mathematical creativity is widely acknowledged. The initial research was an examination of some characteristics of the creative process and the creative person. On the basis of this background, six criteria describing observable aspects of mathematical creativity were identified. These criteria were face validated by seven Professors of Mathematics at the University of Wisconsin and serve as part of a test instrument to measure observable mathematical creativity. One set of conditions conducive to mathematical creativity was proposed and activities which satisfy these conditions were piloted. From these activities both an instructional program to encourage individual mathematical creativity in first grade students and two problems to use a part of the test instrument were developed. An experiment was conducted to determine the effects of participation in the program on observable mathematical creativity; these effects were measured using the test instrument developed for this thesis. The effects on general creativity were measured using the Torrance Tests of Creative Thinking, Figural Forms A and B. The major contributions of this thesis are the identification and face validation of six criteria which describe observable aspects of mathematical creativity and the presentation of evidence that under suitable conditions first grade students can exhibit behaviors satisfying these criteria.

Chapter V

THE TEST INSTRUMENT

5.1 OUTLINE OF CHAPTER V

A test instrument was constructed to measure observable mathematical creativity. The instrument consists in part of six criteria, in terms of observable behaviors, which can be used by observers to evaluate the actions of a person while he is working on a problem. The other part of the instrument is the particular problem on which the person being observed works. The scoring of the pretests and posttests from the experiment was done by a group of five persons, each scoring all pretest and posttests.

The development and face validation of the criteria are discussed first. Then the problems used for pretest and posttest are described. The last section of this chapter reports the selection and training of the scorers, the extent to which the training produced interscorer agreement.

5.2 THE CRITERIA

The author developed the criteria while trying to accomplish a seemingly easier task, that of categorizing a list of many possible responses to a particular problem into two sets: "creative" and "not creative." The problem of consistency in categorization kept arising. If one action was creative, then why was a very similar one not? Was it because one was more clever, or more mathematically relevant, or more unexpected?

Through a process of writing in general terms the essential characteristics of the specific actions which the author considered creative, then returning to the list of specific actions to determine which other items on the list satisfied the general terms, rewriting the general terms and starting all over again, a list of seven criteria were developed. Six of the criteria described actions which might be observed while a person was working on a problem; the seventh criterion pertained to the result of those actions.

The criteria were written in terms of observable behaviors for two reasons. First, it was desired to avoid the problems which can arise if one attempts to infer mental processes, such as "understanding," from actions. It is philosophically and practically more suitable to describe what actions one would accept as demonstrating an aspect of "understanding." The second reason is that a test instrument, to be useful, must be reliable. That is, if several people use the same instrument to measure the same thing, in this case certain aspects of behavior, the measurements should agree quite closely. One of the assets which the author assumed to be true about criteria written in terms of observable behaviors is that such criteria would be used more reliably than criteria requiring that the observer infer mental processes. This assumption has also been made by the American Association for the Advancement of Science, as explained in one of their publications describing <u>Science--A Process Approach</u> (American Association for the Advancement of Science 1965, pp. 1-2).

The author developed the criteria on the basis of background knowledge of the creative process and some personal experience as a mathematician. In order to determine whether the seven criteria really do describe aspects of creative mathematical activity, that is whether they have face validity, the aid of seven Professors of Mathematics at the University of Wisconsin was enlisted. It was decided that each professor should score each of the seven proposed criteria and several dummy criteria on a three point scale: YES (any actions satisfying this criterion would always be an aspect of creative mathematical activity), NO (actions satisfying this criterion would not be an aspect of creative mathematical activity), or MAYBE (unsure, perhaps actions satisfying this criterion would be an aspect of creative mathematical activity). The author would then weigh the responses as 1 point for a YES, 0.5 for a MAYBE, and 0 for a NO. Any criterion or part thereof receiving five or more points, out of a possible seven, was to be accepted as having been face validated.

As the first step in the face validation, the seven professors viewed a thirty minute videotape made from the responses of three of the students videotaped on March 4, 1969. The students were each working individually in the presence of the author on a problem requiring that they trace a triangle as many times as possible on a piece of paper so that the triangles could be cut out. The tapes were edited, but only in those places where the student was continuing the same actions over a long period of time. The professors were then given an earlier form of Chapter I of this thesis to read and appointments were made for the author to interview each one personally within the week. They each agreed not to discuss the criteria until all of them had been interviewed in order that their judgments be independent.

At the interview, each professor was presented with a list of thirteen "criteria" among which were placed, using a table of random numbers, the seven "real criteria" of the author and six "dummy criteria." Each proposed criterion was followed by typical examples from the triangle task; whenever possible, the examples were actions from the videotape. Appended to the list was a glossary defining the way in which certain terms were used. The professor read and discussed each proposed criterion with the author and then scored it, as a whole or in parts, as YES, MAYBE, or NO. The materials used for these interviews and the scores of each proposed criterion are in Appendix B; the results of the face validation proceedings are summarized in the next paragraph and the criteria which were approved follow.

The results of the face validation proceedings were that each of the author's proposed criteria was accepted at a level of 6.5 or seven points out of a possible seven points, with five points having been previously chosen as the minimum level of acceptance. The scores of the dummy criteria ranged from 0 points to 4.5 points; all the dummy criteria were rejected.

The criteria in the form in which they were judged to accurately describe some aspects of observable mathematical creativity are listed here (examples and glossary pertaining to these criteria are in Appendix B):

After the task has been outlined by the teacher, during pursuit of the activity the student:

1. In the absence of a specific stated mathematical goal, verbally introduces some appropriate goal; and/or exhibits goal-directed behavior with respect to some appropriate goal.

2. States an appropriate unstated property of the activity or its product.

3. Conjectures, states, or demonstrates a possible relationship between some appropriate property of the activity and/or products of the current task and some appropriate property of the activity and/or products of either that same task or some previous task; and/or investigates a relationship of the above type.

4. Conjectures, states, or demonstrates a possible generalization; and/or attempts to generalize.

5. Achieves, states, or demonstrates appropriate mathematically elegant product or result.

After the student has pursued the task as outlined by the teacher, the student:

6. Verbally suggests an appropriate modification of the task; and/or exhibits goal-directed behavior with respect to an appropriate modification of the task; and/or conjectures, states, demonstrates, or investigates a possible relationship between some appropriate property of the activity and/or products of an appropriate modifica-

tion of the task and some appropriate property of the activity and/or products of either that same task or some previous task.

7. Verbally suggests an appropriate extension of the task; and/or exhibits goal-directed behavior with respect to an appropriate extension of the task; and/or conjectures, states, demonstrates, or investigates a possible relationship between some appropriate property of the activity and/or products of an appropriate extension of the task and some appropriate property of the activity and/or products of either that same task or some previous task.

The fifth criterion concerns the nature of the result of an activity; the other six criteria describe actions which may or may not lead to a worthwhile result, but which in themselves are aspects of mathematically creative activity because they describe activities which are a normal part of the preparation and manipulation stages in a mathematician's work. The first criterion, setting a goal for oneself, is partly a recognition of the motivational forces involved in the creative process as well as a description of some preparation and manipulation activities. The second, third, and fourth criteria involve the observable aspects of the process by which one notices a mathematical property of something, seeks a relationship between the values of two mathematical properties in a specific case, and seeks a general setting in which a value of some property or a particular relationship exists. The sixth and seventh criteria outline a process by which one might try to use the new idea in a familiar context or explore the further possibilities of the new idea.

The seven Professors of Mathematics expressed the opinion that the distinction which had been made by the author between a modification of a task and an extension of a task was an unnecessary refinement. Consequently the definition of "modification" was rewritten and the two criteria combined, giving a total of six criteria, as reported earlier in this chapter. For the purpose of using the criteria to evaluate observed behavior, some of the criteria were compactified, eliminating the hard to read aspects of the expanded form that seemed necessary in order that the Professors of Mathematics could easily accept part of a criterion and reject other parts. The rewritten criteria are presented in the fourth section of this chapter, SCORING THE VIDEOTAPES.

The criteria are the fixed aspect of the test instrument. The mathematical problem on which those persons being scored according to the criteria would be working can be chosen to suit the purposes of an experiment. The problems used for the pretest and posttest are described in the next section.

5.3 THE MATHEMATICS PROBLEMS USED AS PRETEST AND POSTTEST

Two problem situations, one to use as a pretest and one as a posttest during the experiment, were developed to satisfy certain requirements. The first requirement was that each problem had to satisfy all the conditions placed on the activities in the instructional program.

Other requirements were imposed on the problems by the choice of medium through which the scorers were to do the observing--videotape.

It was decided to videotape all of the pretest and posttest sessions, with each student working individually on a problem in the presence of the author, rather than subject the student to several additional persons who would not only watch, but score his actions. In practice, the students, after at most an initial question about the equipment, seemed to ignore the camera and to become fairly engrossed in the problem. The use of videotape also allowed the author to present the actions of the students to the scorers in one minute intervals, with time for scoring between minutes. The scoring procedure will be discussed more fully in the next section of this chapter.

Since the students' actions while working on the problems were to be videotaped, factors such as the cost of taping, the kinds of student actions evoked by the problem, and the dimensionality of the result of those actions had to be considered. In order to minimize expenses the decision was made to limit each pretest and posttest session to twenty minutes. This meant that the problems chosen had to be ones on which a first grade student could made reasonable progress in that time. Since a videotape camera can focus on small movements or large ones, but a single camera cannot do both at the same time or instantaneously change from one type to the other, the problem chosen must not generate important student actions of both a large and small scope in a frequently alternating sequence.

The above two considerations were not very difficult to satisfy; the dimensionality consideration was quite a bit more restricting. One problem suitable in many requirements but not the one of dimension is the problem of how many triangles one could make with six straws because some of the solutions or partial solutions may involve three dimensional structures. It is sometimes difficult to correctly visualize a three dimensional object from the two dimension views of it given on a television screen, and consequently evaluating the actions which produced such an object, and the object itself, according to the six criteria would be difficult and prone to error. It was decided to restrict the pretest and posttest problems to situations which are planar in their essential details.

One acceptable problem was piloted with ten students at the Prospect Street Elementary School, Lake Mills, Wisconsin, on March 4, 1969. This was the problem of tracing an equilateral triangle as many times as possible on a sheet of paper so that the triangles could be cut out. As each student was working on the problem, his actions were videotaped so that videotapes similar to those planned for use in the experimental testing sessions could be available for several purposes. The two anticipated purposes which the videotapes served were as a means of training the persons who acted as scorers for the tapes of the experimental testing sessions, and as a basis for a thirty minute edited tape which could introduce people to the work of the author. The latter of these purposes was mentioned in the previous section of this chapter; the former will be discussed more fully in the next section.

Two unanticipated kinds of information were provided by the videotapes. It became obvious from watching the tapes that materials for the tests had to be very carefully chosen if the goal of seeing clearly what was taped was to be reached. Of great importance to the testing sessions in the experiment was the discovery that during the March 4 videotaping the author unknowingly had responded inconsistently to all the students. As a result, the problems used in the experiment were written with responses indicated and a pattern of steps were carefully spelled out in order to insure as much consistency as possible.

Each of the problems used for testing is briefly described here. A more complete description of the problems, including the rules used by the author to aid consistency, is in Appendix C.

The materials for the pretest problem were white 5 1/2" x 8 1/2" paper, a cardboard triangle 2" per side, a black felt-tipped marker, a scissors, and a cardboard square and a cardboard diamond 2" per side each.

The task was posed to the student as follows: The author ascertained that the student knew the name "triangle." Then she asked, "How many times do you think you could trace this triangle onto this piece of paper so that you could cut the triangles out?"

The materials for the posttest problem were plastic trapezoid tiles of the shape shown in (a) of Figure 5.1 and three each of white cardboard forms with black indentations in one of seven shapes:

(a) circle, diameter 2 1/16"

(b) square, side 2 1/16"

(c) trapezoid, shape (b) of Figure 5.1

(d) trapezoid, shape (c) of Figure 5.1

(e) regular hexagon, side 1 3/8"

(f) equilateral triangle, side 2 1/16"

(g) diamond, side 2 1/16".

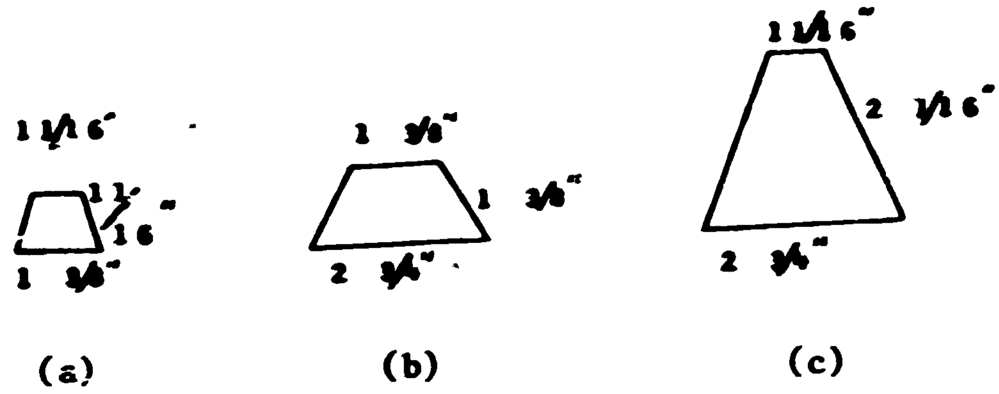


Figure 5.1

Three Trapezoids Used in the Posttest Problem

The task was posed to the student as follows: "Some of these shapes can be filled in with these tiles and some can't. Can you find out which ones can by filling them up and show me why some of the shapes can't be filled up using these tiles?"

Both problems require movements of small scope and are planar. The experience of the author has been that each can be completed by the average first grade student to his satisfaction in less than twenty minutes; often the student considered that he was done in less than ten minutes, especially on the trapezoid tile problem.

Both problems involve the kinds of juxtapositions possible with many copies of one planar shape, in one task an equilateral triangle, in the other a quadrilateral. This content is shared with the activities of the instructional program. Both tasks satisfy the requirements set for the instructional activities in terms of content, concrete embodiment of the problem, open-endedness, and moderate amount of structure.

The first activity of the instructional program, a tiling activity, could result in a tessallation of a planar section by triangles. Because it was desired to measure, on the pretest and posttest, primarily the actions of the student in approaching a new problem, not just his result, it was decided that the triangle task should be the pretest, not the posttest. The trapezoid shape was not used as part of the instructional materials.

5.4 SCORING THE VIDEOTAPES

Five graduate students at the University of Wisconsin were employed to serve as scorers; three were mathematics graduate students, one was a psychology graduate student, and one was an educational psychology graduate student. The distribution of scorers were deliberate, the assumption being that the mathematics students might interpret mathematical relevance differently than the psychology and educational psychology students. It was expected that the differences in academic discipline might also cause a difference in the scoring of the non-verbal actions.

The students were interviewed by the author previous to the beginning of the training sessions. It was considered preferable but not necessary that the mathematics students have had some experience with children of about first grade age. Two of the three satisfied this preference. Each of the non-mathematics students satisfied the author during his interview that mathematics did not arouse negative reactions in him. It was felt that a person who could not view the mathematical aspects of the testing problems as interesting would not be a good scorer.

The Scoring Procedures

Before the first training session, each scorer received and read a description of the scoring procedure which included a sketch of each of the testing problems, a list of the criteria with a glossary, and a sample scoring sheet. These materials are in Appendix D.

The criteria and some of the definitions were rewritten after the face validation to make them more compact. The expanded form was useful to the Professors of Mathematics for the purpose of judging whether one wished to accept only part of a criterion, but was more clumsy for scoring purposes. As part of the compactification two criteria were combined. The rewritten criteria are listed here; the relevant glossary is in Appendix D. It is the belief of the author that the rewriting was consonant with the criteria as approved by the Professors of Mathematics and with the comments made by them.

1. Introducing a goal: In the absence of a specific stated mathematical goal, the student either verbally suggests or exhibits goal-directed behavior with respect to some appropriate goal.

2. Identifying a property: The student states an appropriate unstated property of the task.

3. Seeking a relationship: The student conjectures, states, demonstrates, or investigates a possible relationship between (a) some

appropriate property of the task he is pursuing and (b) either some other appropriate property of the same task or some appropriate property of some different task.

4. Seeking a generalization: The student conjectures, states, demonstrates, or attempts a possible appropriate generalization.

5. Reaching a mathematically elegant product: The student achieves, states, or demonstrates a mathematically elegant and appropriate product.

6. Modifying the task: After having pursued the task as outlined by the teacher, the student either verbally suggests or pursues an appropriate unstated modification of the task.

The scoring was done by viewing a tape for one minute, then each scorer individually and silently scoring the actions of the student during that minute on each of the six criteria. It was felt that the scorers could remember the actions from one minute and that a smaller time interval would not greatly add to the power of the test instrument. Each criterion was to be considered independently of the others in the sense that each could be satisfied or fail to be satisfied regardless of whether another criterion was satisfied during that minute. A criterion was to be marked as satisfied during a particular minute if one or more sets of actions satisfied that criterion, with no consideration given to whether the student had exhibited some of those actions during previous minutes. The training sessions were conducted like scoring sessions except that after each minute was scored, a discussion of the scoring preceded the viewing of the next minute.

The nature of the actions on the training tapes, those made on March 4, 1969, of the students trying to maximize the number of triangles which could be traced on a piece of paper, seemed to make it difficult for the scorers to score according to the guidelines just listed. When several minutes passed during which nothing happened worth scoring, and then in one minute the student investigated two relationships, the scorers became disconcerted that the student was not getting "credit" for both. There was a great reluctance on the part of some scorers to mark a criterion satisfied minute after minute when the student was repeating the same actions which they had scored previously. Also, a tendency developed to score anything and everything as "relationship."

The author tried several variations of the scoring procedure in order to overcome these problems. First, the author shared with the scorers her own misgivings about occassionally "losing information" because of the way the minutes fell, but attempted to explain to them that most measurement devices lose information of some sort, and that this was the price one paid for trying to reduce behavior to numbers. It was suggested that if it made the scorers feel better, they could put two check marks in the box pertaining to one criterion and one minute, but that the author would count them as one. This seemed to satisfy the scorers and the problem disappeared.

In order to encourage continued scoring of continued actions which satisfy some criterion, a refinement was added. The first time some action occurred, it was scored with an "N" for new; any repetitions were scored "O" for old. Again this device was for the benefit of the scorers only; the author made it clear to them that she would not distinguish between these scores. This procedure worked reasonably well, although it did raise the question of whether or not one could score one criterion both "N" and "O" in the same minute.

Further complicating the scoring was an attempt to force the scorers to decide more carefully which criterion or criteria were satisfied by a set of actions. The tendency to score everything as "relationship" probably resulted from the fact that most of the scorable actions were "relationship," so scoring actions under that heading became like a habit. To counteract this, a two pass scoring procedure was tried. The tape was to be viewed in its entirety twice; the first time the scorers were to score whether or not "something" had happened, the second time they were to decide what it was. This was a complete failure. The scorers found that in trying to decide if "something" happened, they were actually examining the actions of the student in the light of each criterion; this made the second pass superfluous and boring.

As a result of these attempts, almost eleven hours worth of training time, two things happened. The scorers were so tired of refinements, they were willing to drop the "N" and "O" classification. As far as the author was concerned, that was good. The other thing was that by this time all the training tapes had been seen by the scorers, several of them twice, so that the only fresh tapes which could be used to calculate interscorer agreement were the tapes from the experiment.

The author felt that agreement on the scoring of the pretest task would be reasonable, since all of the training tapes had been of that task, but both the author and the scorers felt that there needed to be more exposure to the kinds of actions with which a first grade student might respond to the trapezoid task. No videotapes of students working on that task were available, so a live acting situation was set up. Miss Joan Moscovitch, an employee of the Wisconsin Research and Development Center, who has had several years experience teaching in the early primary grades, volunteered to act the part of a first grade student in a live demonstration for the scorers. The author set up a table as if the situation were a posttest and Miss Moscovitch and the author acted out a typical posttest session, stopping every minute or so for the scorers to discuss what they had seen. This procedure was followed for about half an hour, at the end of which time the scorers felt that they were sufficiently prepared for the actual scoring of this task.

To avoid problems of bias in the scores, the author did not tell the scorers which of the tasks was the pretest and which the posttest. There was no indication given on the tapes or by the author as to which students were in the program. As a further precaution, the order in which the tapes were scored was chosen at random. Each testing session yielded two reels of tape; they were scored in the following order: the second pretest tape, the first posttest tape, the second posttest tape, the first pretest tape.

The Interscorer Agreement

In order to calculate interscorer agreement, the author tried to follow Winer (1962, pp. 24-28). Unfortunately, the statistic he offers for calculating interjudge reliability was not suitable, because it presumes continuous data and the scores in this case were dichotomous. There seemed to be no other statistic available which would measure interscorer agreement. After consultation with Mr. Thomas Fischbach, a statistician employed by the Wisconsin Research and Development Center for Cognitive Learning, the author settled on the following method.

Under the assumption that the majority of the scorers must be correct, any scorer who deviates from the majority opinion must be in error. With five scorers, the situation could be total agreement, or one of two splits, a four-one or a three-two. Thus it would be possible to calculate for each scorer on the application of each criterion a rate of error. This was done, breaking the errors into two kinds: not scoring a criterion as satisfied when the majority did, called "error of omission," and scoring the criterion as satisfied when the majority did not, called "error of commission." Both of these partial error rates, plus their sum, the combined error rate in which the error accounted for is the scoring of a criterion differently from the majority, were calculated for each of the pretest, posttest, and their average, called "overall." In order to help interpret these error rates, the proportion of minutes in which each criterion was considered satisfied by a majority of scorers was calculated for each task and overall.

in the case of five scorers, a three-two split is total disagreement. Following the meanings given to correlation coefficients, one could assign a "-1" to every minute during which a criterion was scored by a three-two split, a "+1" to every minute during which a criterion was scored by a five-zero "split." The choice of assigning a "0" to those minutes during which there was a four-one split does not really fit the model by signifying no agreement, but with only three scoring situations possible, it seemed reasonable to assign a "0" to those events. Using this method, an interscorer agreement was calculated on each of the six criteria for each task and overall.

In order to examine whether the scorers agreed on the existence of something in a minute which was worth scoring, even if they disagreed on what criterion it satisfied, a factor called "any" was included as a pseudocriterion. This pseudocriterion was considered satisfied if any one of the six criteria was satisfied, and the same statistics calculated for it as were calculated for the six criteria.

An abbreviation of the major element in each criterion, rather than numbers, are used in all the tables to refer to the six criteria. The abbreviations are "goal" for criterion 1; "prop" for property, criterion 2; "rel" for relationship, criterion 3; "gen" for generalization, criterion 4; "prod" for product, criterion 5; and "mod" for modification, criterion 6. The pseudocriterion is abbriviated as "any."

In examining the error rates of the scorers and the interscorer agreements, the fact that some criteria were never or rarely considered as satisfied by a majority of the scorers must be taken into account, since for these criteria, the error rates are low and the interscorer agreement is high. If the numerical interpretation of "never or rarely considered as satisfied" is set as satisfied during less than 5 percent of the minutes, then for the pretest, criteria 1, 2, 4, 5, and 6 fall into this classification; for the posttest criteria 1, 4, and 6 were so classified; and overall, criteria 1, 4, and 6, are again in this classification. This information is summarized in TABLE 5.1.

TABLE 5.1

CRITERION CONSIDERED SATISFIED BY A MAJORITY OF THE SCORERS

Task	any	goal	prop	rel	gen	prod	mod
Pretest	.14	.00	.00	.13	.03	.01	.00
Posttest	.90	.00	.15	.78	.00	.39	.00
Overall	.52	.00	.08	.46	.01	.20	.00

Reported as the proportion of minutes.

In the discussion of the error rates for the scorers on the two tasks and overall, there will be few comments made about the error rates on the criteria satisfied during less than five percent of the minutes because there is not enough data on the scoring of these criteria to make such comments very meaningful. However all the criteria will be discussed with respect to interscorer agreement because agreement that a certain criterion was not satisfied is as important to the use of the test instrument as agreement that the criterion was satisfied.

Making either one of the particular errors during more than 7 percent of the minutes will be considered a high rate of error; erring in the sum of both ways during more than 10 percent of the minutes will be considered a high rate of error. Error rates lower than these will be considered as low. Interscorer agreements .80 or higher will be considered as high, between .50 and .80 as reasonable, and lower than .50 as low.

In the pretest only the pseudocriterion and criterion 3 were scored as being satisfied with any reasonable frequency, and both of those were considered as satisfied during less than 15 percent of the 88 minutes of pretest tape. The error omission was committed by no scorer during more than 5 percent of the minutes. The error of commission was committed by scorer three during at least 15 percent of the minutes. The combined error is high, 18 percent, for scorer three, and low, less than 10 percent, for the other four scorers. Since only one criterion was scored as frequently satisfied during the pretest, the scoring of pseudocriterion tends to reflect the scoring of that criterion. These data are summarized in TABLE 5.2 on page 102.

In the posttest, the pseudocriterion and criteria 1, 4, and 6 were scored as satisfied with reasonable frequency during the 96 minutes taped. No scorer made errors of commission during more than 7 percent of the minutes. However, scorer three made errors of omission during 28 percent of the minutes. This is also reflected in a similar error in his scoring of the pseudocriterion during 23 percent of the minutes. Scorer four also made this error on criterion 3 fairly often, during 14 percent of the minutes. On the other criteria or for the

TABLE 5.2

PRETEST: ERROR RATES OF THE SCORERS

Scorer	any	goal	prop	rel	gen	prod	mod
			Error of Omission				
1	.01	.00	.00	.05	.00	.00	.00
2	.01	.00	.00	.02	.00	.00	.00
3	.02	.00	.00	.03	.00	.00	.00
4	.00	.00	.00	.00	.01	.00	.00
5	.00	.00	.00	.01	.02	.00	.00
			Error of Commission				
1	.00	.00	.00	.00	.00	.00	.00
2	.05	.00	.00	.06	.02	.00	.00
3	.16	.00	.00	.15	.02	.00	.00
4	.05	.00	.01	.03	.01	.01	.00
5	.03	.00	.02	.00	.01	.02	.00
			Combined Error				
1	.01	.00	.00	.05	.00	.00	.00
2	.06	.00	.00	.08	.02	.00	.00
3	.18	.00	.00	.18	.02	.00	.00
4	.05	.00	.01	.03	.02	.01	.00
5	.03	.00	.02	.01	.03	.02	.00

Reported as the proportion of minutes in which the error was made.

other scorers, this error was made during no more than seven percent of the minutes; the combined error rate was low—under 10 percent—except in the same cases as for the error of omission: scorers three and four on criterion 3 and scorer three on the pseudocriterion. This information is summarized in TABLE 5.3 on page 104.

The scoring of both tasks, the overall scoring, was similar to the scoring of each of the individual tasks in that for each of the particular errors and the combined error, on most of the criteria and for most of the scorers, particular errors were made during no more than 7 percent of the minutes and the combined error was made during no more than 10 percent of the minutes. The only exception is scorer three on criterion 3 and the pseudocriterion, both on the error of omission and the error of commission. Combining the errors shows scorer three committing an error on criterion 3 during 25 percent of the minutes and on the pseudocriterion during 21 percent of the minutes. This certainly leads to the speculation that low overall interscorer agreements on both the pseudocriterion and criterion 3 might be caused by the high error rate of scorer three on criterion 3 which is also reflected in his high error rate on the pseudocriterion. These data are summarized in TABLE 5.4 on page 105.

Interscorer agreements were calculated on each of the six criteria and the one pseudocriterion for each task and overall. In addition, the average of the interscorer agreements on the six criteria, a total interscorer agreement, was calculated; it is abbreviated as "ave."

TABLE 5.3

POSTTEST: ERROR RATES OF THE SCORERS

Scorer	any	goal	prop	rel	gen	prod	mod
			Error of Omission				
1	.00	.00	.01	.00	.00	.00	.00
2	.02	.00	.00	.04	.00	.00	.00
3	.23	.00	.00	.28	.00	.02	.00
4	.06	.00	.01	.14	.00	.01	.00
5	.00	.00	.01	.03	.00	.04	.00
			Error of Commission				
1	.04	.00	.01	.07	.00	.03	.00
2	.01	.00	.00	.04	.00	.02	.00
3	.01	.00	.02	.04	.00	.00	.00
4	.00	.00	.03	.00	.00	.00	.00
5	.02	.02	.02	.03	.01	.01	.01
			Combined Error				
1	.04	.00	.02	.07	.00	.03	.00
2	.03	.00	.00	.08	.00	.02	.00
3	.24	.00	.02	.32	.00	.02	.00
4	.06	.00	.04	.14	.00	.01	.00
5	.02	.02	.03	.06	.01	.05	.01

Reported as the proportion of minutes in which the error was made.

TABLE 5.4

OVERALL: ERROR RATES OF THE SCORERS

Scorer	any	goal	prop	rel	gen	prod	mod
			Error of Omission				
1	<.01	.00	<.01	.03	.00	.00	.00
2	.015	.00	.00	.025	.00	.00	.00
3	.125	.00	.00	.155	.00	.01	.00
4	.03	.00	<.01	.07	<.01	<.01	.00
5	.00	.00	<.01	.02	.01	.02	.00
			Error of Commission				
1	.02	.00	<.01	.035	.00	.015	.00
2	.03	.00	.00	.05	.01	.01	.00
3	.035	.00	.01	.095	.01	.00	.00
4	.025	.00	.02	.015	<.01	<.01	.00
5	.025	.01	.02	.015	.01	.015	<.01
			Combined Error				
1	.025	.00	.01	.06	.00	.015	.00
2	.045	.00	.00	.08	.01	.01	.00
3	.21	.00	.01	.25	.01	.01	.00
4	.055	.00	.025	.085	.01	.01	.00
5	.025	.01	.025	.025	.02	.035	<.01

Reported as the proportion of minutes in which the error was made.

In the scoring of the pretest, the total interscorer agreement is quite high, .92. This is somewhat deceptive in that very little was scored at all, reflecting the fact that, in the opinion of the author, during many of the minutes of the pretest tapes nothing happened which could be considered for scoring, so that there was little source of disagreement on the tapes. Those few actions which were considered for scoring fell almost exclusively under criterion 3, "relationship." The interscorer agreement is .63 for both criterion 3 and the pseudocriterion. The posttest scoring also has a high total interscorer agreement, .83, due in part to the higher agreements on the criteria less frequently considered satisfied. The agreement on criterion 3 is disturbingly low, .26, especially because the fact that a criterion or pseudocriterion is frequently considered as satisfied should not mean that interscorer agreement is low. The pseudocriterion was considered as satisfied during 90 percent of the minutes of the post test, as compared with 78 percent for criterion 3, and the interscorer agreement on the pseudocriterion is a reasonable .59. Criterion 5 was considered as satisfied during 39 percent of the minutes of the posttest with an interscorer agreement of .85. Interscorer agreement on criterion 2 is .89. In the overall scoring, the total agreement is high, .87, again partially attributable to the higher agreements on the criteria less frequently considered satisfied. The interscorer agreement on the pseudocriterion is a reasonable .61 and on criterion 3 a low .43. These data are summarized in TABLE 5.5.

TABLE 5.5

INTERSCORER AGREEMENT

Task	any	goal	prop	rel	gen	prod	mod	ave
Pretest	.63	1.00	.98	.63	.94	.96	1.00	.92
Posttest	.59	1.00	.89	.26	.98	.85	.98	.83
Overall	.61	1.00	.93	.43	.96	.90	.99	.87

Because scorer three had a combined overall error rate of 25 percent on criterion 3, the author suspected that some of the lower interscorer reliabilities were in part caused by this one scorer, and decided to calculate an interscorer agreement for the other four scorers. This was done by assigning a +1 to a four-zero split, a zero for a three-one split, and a -1 for a two-two split; these calculations were made for pretest, posttest, and overall. For the pretest, the partial interscorer agreement on criterion 3 is a high .81 and on the pseudocriterion it is a high .80, as compared with a reasonable .63 for each of them when calculated for all five scorers. For the posttest, on criterion 3, the partial interscorer agreement becomes a reasonable .62, as compared to a low .26; and on the pseudocriterion, the reasonable .59 becomes a high .97. On criteria 2 and 5, almost no change occurs--.89 becomes .90 and .85 becomes .88 respectively. The interscorer agreements for the overall scoring show similar effects to those of the posttest after the removal of the one scorer. The agreement on the pseudocriterion rises from .61 to .89; on criterion 3, from .43 to .71; on criterion 2, from .93 to .94; and on criterion 5, from .90 to .93. The partial interscorer agreements are given in TABLE 5.6.

TABLE 5.6

INTERSCORER AGREEMENT OF SCORERS 1, 2, 4, AND 5

Task	any	goal	prop	rel	gen	prod	mod	ave
Pretest	.80	1.00	.98	.81	.94	.98	1.00	.95
Posttest	.97	1.00	.90	.62	.99	.88	.97	.89
Overall	.89	1.00	.94	.71	.97	.93	.99	.92

The increases in interscorer agreement produced by eliminating scorer three from the calculations are dramatic in the case of criterion 3 and the pseudocriterion. In a conversation with scorer three after the scoring was finished, the author learned that this scorer was consciously applying criterion 3 using his own guidelines, rather than the ones which were supposedly agreed upon by all the scorers. He said he did this because he could not explicitly recall those group guidelines. This would explain the differences between his scoring and that of the other scorers, but it does pose some interesting questions about the use of the test instrument. These questions and others are discussed in Chapter VII, CONCLUSIONS.

The experiment which was conducted to test the instructional program discussed in Chapter IV using the test instrument discussed in this chapter is described in the next chapter, Chapter VI.

Chapter VI

THE EXPERIMENT

6.1 OUTLINE OF CHAPTER VI

An experiment was conducted to test whether the program developed to present first grade students with open-ended mathematical problems would increase the observable mathematical creativity of those students who participate in the program. A review of the specifics of the problem to be examined in the experiment is given in the next section of this chapter. The design of the experiment and a discussion of the statistical analyses suggested by that design is in the following section. The last section of the chapter contains the data from the experiment.

6.2 RESTATEMENT OF THE PROBLEM

The problem examined in the experiment involves the effects of a treatment in the form of an instructional program consisting of a sequence of open-ended mathematical problems developed to encourage the observable mathematical creativity of first grade students. The program used in the experiment is described in Chapter IV of this thesis; Chapters II and III present the theoretical aspects of the development of the program. The general hypotheses to be tested were

H1: Participation in the program will increase a student's observable mathematical creativity; and H2: Participation in the program will not affect a student's performance on a test of general creative ability. Hypothesis H1 is stated in terms of an expected increase because the program is designed to encourage a student's mathematical creativity. Because no attempt is made in the program to encourage general creativity, and because transfer from the techniques and attitudes mathematical creativity to those of general creativity is not likely unless a student on his own perceives the two kinds of situations as similar, hypothesis H2 is stated in terms of no expected change.

The test instrument used to measure observable mathematical creativity is the one developed by the author and described in Chapter V of this thesis. Each student is given credit for satisfying a criterion during a particular minute if at least four of the five scorers considered that criterion as satisfied during that minute. For each student on both the pretest task and the posttest task, a vector of eight number is obtained from this test instrument: the number of minutes the student is observed and scored via videotape; the number of minutes the student satisfied at least one criterion; and for each of the six criteria, the number of minutes the student satisfied it. In some cases, technical problems with the videotape equipment caused a student to be scored on fewer minutes than he actually worked. The effects of these problems are discussed more fully later.

The Torrance Tests of Creative Thinking, Figural Forms A and B, were used as the instrument for measuring general creative ability.

These tests yield four scores--fluency, flexibility, originality, and elaboration--but only three were used; elaboration was eliminated because it seemed to be the most difficult dimension to score reliably. The scoring manuals for these tests give the flexibility categories and originality weights to be used in scoring as well as guidelines for scoring responses not listed in the manual. Reliability of the scoring procedure between trained scorers and those who have been introduced to the scoring procedures only by careful reading of the manuals is "rather consistently above .90" (Torrance 1966f, p. 18). It was assumed that identical raw scores on the two forms of the Figural Tests would not necessarily have the same meaning, and that conversion of the raw scores to standard (T) scores would improve comparison of pretest and posttest scores because the same meaning would be attached to a particular score regardless of the form from which it was obtained. The only available conversion tables were based on fifth-grade students (Torrance 1966f, pp. 61 and 66). However, Torrance has found that the T scores based on this fifth-grade data "lend themselves satisfactorily to conversions at both the lower and upper levels educationally" (Torrance 1966f, p. 57).

Of the three sets of parallel activities in the Torrance Tests of Creative Thinking, Figural Forms A and B, the first set was the only one which could not be scored on the fluency dimension. In order to shorten the time required to complete the test, the first activity was omitted during both the pretest and posttest sessions.

6.3 THE EXPERIMENTAL DESIGN

The relationship of the mathematical problems used as the pretest and posttest to the problems in the instructional program--the former were required to be similar to the latter in some elements of both structure and content--suggested that an experimental design be chosen which could control for and measure the effect of participation in a pretest. One design which can control for and measure this effect is the Solomon Four-Group Design (Campbell and Stanley 1966, pp. 24-25).

The Details of the Experimental Design

Subjects, all members of the same first-grade classroom of twenty-seven students, were randomly assigned to one of four experimental groups from a stratified sample based on mathematics achievement. A stratified sample was used for two reasons. First, because general creativity is found distributed among all levels of IQ and scholastic achievement (Guilford 1962, pp. 163-4), it is possible that mathematical creativity is also distributed among all levels. Second, the nature of the creative process seems to indicate that trying to encourage mathematical creativity in the child who has demonstrated less than average mathematical achievement may provide that child with a rewarding mathematical experience which would invigorate his interest in the subject. A table of random numbers was used to select one student for each of the four experimental groups from each of the three strata of nine students. This gave each experimental group a population of three; thus twelve of the twenty-seven students in the class were involved in the experiment.

Group 1 received pretest, treatment, and posttest. Group 2 received pretest and posttest, no treatment. Group 3 received treatment and posttest, no pretest. Group 4 received posttest only, no pretest and no treatment. Groups 1 and 3 received the same treatment at the same time as a group of six. This design is summarized in TABLE 6.1

TABLE 6.1

SOLOMON FOUR-GROUP DESIGN

Group	Assignment	Pretest	Treatment	Posttest
1	RS	yes	yes	yes
2	RS	yes	no	yes
3	RS	no	yes	yes
4	RS	no	no	yes

RS indicates random assignment to groups from a stratified sample. This TABLE is adapted from Campbell and Stanley (1966, p. 24).

The pretesting and posttesting each involved two tests. One was the test designed by the author, the other was the Torrance Tests of Creative Thinking, Figural Form A for posttest and Form B for pretest.

The pretesting and posttesting were each done in one day. On the pretesting day, the mathematical creativity test was given in the morning, with the order in which the students took the test decided by the use of a table of random numbers; the general creativity test was given to all six students as a group in the afternoon. On the post-testing day, the general creativity test was given in the morning to two groups of six students, one of the groups consisting of those students who were pretested; the mathematical creativity test was given starting in mid morning and ending in the early afternoon, with the order of the students again decided by use of a table of random numbers. Scheduling problems involving the videotape equipment were the reason that the same morning and afternoon times were not used for the two types of tests on both the pretesting and posttesting days. All students were present on both days.

Statistical Analyses Appropriate to the Experimental Design

Each student in the experiment is represented by eleven scores for each testing session in which he participated: one score is the number of minutes the student was observed via the videotape, seven scores measure aspects of mathematical creativity; three scores measure aspects of general creativity. These scores for each testing session are given in TABLE 6.2. Abbreviations are used as headings in the table. For ID (identification) a three letter symbol, ABC, is given for individual scores where A = Pretest or No pretest, B = Instruction or No instruction and C = mathematics achievement level: High, Middle, or Low; a two letter symbol, AB, is given for group averages where A and B are as above. For the mathematical creativity scores, minutes videotaped is abbreviated as "min," the pseudocriterion as "any," criterion 1 as "goal," criterion 2 (property) as "prop," criterion 3 (relationship) as "rel," criterion 4 (generalization) as "gen," criterion 5 (product) as "prod," and criterion 6 (modification) as "mod." All these scores are reported as the number of minutes and the averages are to the nearest whole number. For the general scores, fluency is abbreviated as "flu," flexibility as "flex," and originality as "orig." These scores are reported as standard (T) scores and the averages are to the nearest whole number.

TABLE 6.2

INDIVIDUAL SCORES AND GROUP AVERAGES

			Mathematical Creativity						General Creativity		
ID	min	any	1 goal	2 prop	3 rel	4 gen	5 prod	6 mod	flu	flex	orig
				Pretest							
PIH	13	1	0	0	1	0	0	0	34	38	50
PIM	15	2	0	0	2	1	1	0	50	50	52
PIL	14	0	0	0	0	0	0	0	47	43	50
PI	14	1	0	0	1	0	0	0	44	44	51
PNH	18	3	0	0	1	2	0	0	47	48	52
PNM	11	0	0	0	0	0	0	0	40	43	41
PNL	17	2	0	0	2	0	0	0	34	37	40
PN	15	2	0	0	1	1	0	0	41	43	44

TABLE 6.2 (CONTINUED)

INDIVIDUAL SCORES AND GROUP AVERAGES

	Mathematical Creativity								General Creativity		
ID	min	any	goal	prop	rel	gen	prod	mod	flu	flex	orig
				Posttest							
PIH	6	3	0	0	1	0	2	0	60	65	65
PIM	12	11	0	0	10	0	3	0	53	65	65
PIL	3	3	0	0	3	0	1	0	53	53	63
PI	7	6	0	0	5	0	2	0	55	61	64
PNH	7	6	0	2	5	0	3	0	62	58	64
PNM	8	7	0	1	5	0	3	0	49	55	56
PNL	7	4	0	1	3	0	2	0	53	58	65
PN	7	6	0	1	4	0	3	0	55	57	62
NIH	8	6	0	1	5	0	2	0	50	55	56
NIM	11	11	0	1	10	0	5	0	52	45	42
NIL	8	6	0	2	5	0	4	0	52	47	53
NI	9	8	0	1	7	0	4	0	51	49	50
NNH	11	9	0	1	5	0	3	0	63	53	62
NNM	9	8	0	2	8	0	1	0	60	65	84
NNL	7	6	0	2	4	0	4	0	62	55	57
NN	9	8	0	2	6	0	3	0	62	58	68

From TABLE 6.2 it can be seen that of the seven aspects of mathematical creativity, each of which is scored as the number of minutes during which the student satisfied one of the six criteria or the pseudocriterion, three aspects--criteria 1, 2, and 6--are zero for every student in the pretest and three aspects--criteria 1, 4, and 6--are zero for every student in the posttest. Of the aspects which have non-zero scores, only three are non-zero for both the pretest and the posttest; these are criteria 3 and 5, which pertain to seeking relationships and achieving products, respectively, and the pseudocriterion, which is satisfied if at least one of the six criteria is satisfied.

The pseudocriterion and criterion 3 show the greatest variation in scores. This and other considerations discussed later in this section led to the decision to do a complete analysis on only these two aspects and to do a less complete analysis on criteria 2 and 5. Since criteria 1 and 6 were always zero, no analyses were done on them. This leaves criterion 4, which was scored non-zero only by two of the six students in the pretest. Because one of these students was in the treatment group and one in the control group, it seems that no meaningful information would be obtained by analysis of the scores on criterion 4.

The analyses made, following the recommendations of Campbell and Stanley (1966), include a 2 x 2 analysis of variance on the posttest scores examining the effects of having or not having pretest and treatment. For Groups 1 and 2, the difference between posttest scores and pretest scores, usually called gain scores, are examined in two ways: an analysis of variance is made to determine the effect of treatment on the gain scores and an analysis of covariance is made to determine the effect of treatment on the gain scores using the pretest scores as a covariate. In addition, a correlation matrix is computed for the posttest scores on the three aspects of general creativity and the two aspects of mathematical creativity. These analyses are presented in the next section.

6.4 THE DATA

The experiment was designed to test two main hypotheses, one concerning the effect of treatment on the mathematical creativity scores and the other concerning the effect of treatment on the general creativity scores. Each main hypothesis was broken into several sub-hypotheses which in turn are divided into parts and are tested using, in all cases, the method of analysis of variance; for each sub-hypothesis part, a summary of the analysis of variance is presented. The data are grouped under the two general hypotheses; the mathematical creativity scores are discussed first. The correlation matrix is presented after the analyses of the general creativity scores. All of the analyses were made using the REGAN1 program (Guha 1966).

Analyses Performed on the Mathematical Creativity Scores

The general hypothesis concerning mathematical creativity is H1.

H1: Participation in the program will increase a student's observable mathematical creativity.

This general hypothesis is divided into three experimental hypotheses, each of which is examined with respect to each of the relevant mathematical creativity scores.

The first experimental hypothesis is H1.1.

H1.1: There is not a positive effect of pretest, treatment, and their interaction on the number of minutes the student satisfies a criterion or the pseudocriterion when the number of minutes the student was observed is taken as a covariate.

Hypothesis H1.1 has four parts: H1.1a concerns the pseudocriterion, H1.1b concerns criterion 3, H1.1c concerns criterion 2, H1.1d concerns criterion 5.

All parts of hypothesis H1.1 were examined using the following model, developed by Mr. Thomas Fischbach, a statistician at the Wisconsin Research and Development Center for Cognitive Learning. Assume that the number of minutes satisfying a criterion or pseudocriterion is a binomial random variable with each minute of observation an "independent trial." Although whether the actions observed satisfy a criterion does sometimes depend on what preceded those actions, and a reordering of the individual minutes would not necessarily result in the same scores, it is nevertheless reasonable to assume that the probability of a criterion being satisfied during any one minute is not dependent on the order of the minute in the sequence or the scores of the previous minutes. In other words, the scores of previous minutes do not predict the actions of the present minute.

Let y_{ijk} be the number of minutes satisfying a criterion or pseudocriterion and n_{ijk} be the number of minutes observed for the kth subject (k = 1,..., 3) in the ith "pretest condition" and the jth "treatment condition." Then, if u_{ij} is an additive factor depending on pretest and treatment conditions, the familiar formula for the expectation of a binomial random variable becomes

$$\text{Expectation } (y_{ijk}) = u_{ij} + n_{ijk}p_{ij}$$

where p_{ij} is the probability of satisfying the criterion or pseudocriterion in any one minute of observation and p_{ij} depends on pretest and treatment conditions.

There are several possibilities: u_{ij} could be a constant u, that is, the additive factor could be independent of pretest and treatment conditions; p_{ij} could be a constant p (with similar interpretation); both u_{ij} and p_{ij} could be constants; p could be equal to zero.

The mathematical creativity scores are counts of the number of minutes in which satisfying actions occurred. Because they are counts, each would tend to be distributed as a Poisson random variable rather than a normal random variable, and would have a variance dependent on the mean of the distribution. The method of analysis of variance is based on assumptions of normality and equal variances among experimental groups, neither assuption satisfied by these scores. However, the "robustness" of this method makes it a suitable test.

The analyses testing the parts of hypothesis H1.1 each proceeds in two steps. First u and p are fit as constants, using the number of minutes observed (n_{ijk}) as the independent variable. Then in a second step pretest, treatment, and interaction are accounted for by looking at the differences $u_{ijk} - u$ and $p_{ij} - p$. For each step in the fitting process, an F-ratio is calculated as an indication of the significance of the contribution to the sum of squares made by the independent variables added in that step.

This method of analysis shows that for hypothesis part H1.1a, the effect of pretest, treatment and interaction on minutes satisfying the pseudocriterion with minutes observed as a covariate, both u and p are best fit as constants; the values given by regression are $u = -1.5878$ and $p = 1.0$ (subject to the restriction that $p \leq 1$). For the contribution of u and p, which is the contribution made by the covariate minutes observed, the F-ratio is 41.41 which is significant at $p < .005$. The contribution of $u_{ij} - u$ or $p_{ij} - p$, which is the contribution of pretest, treatment and interaction, does not greatly increase the sum of squares predicted by the model; the F-ratio for these additions is 0.41 which is not significant. Thus, hypothesis part H1.1a is not rejected. This analysis is summarized in TABLE 6.3.

Hypothesis part H1.1b, the effect of pretest, treatment and interaction on minutes satisfying criterion 3 with minutes observed taken as covariate, is tested using the same model. The results are similar but not as significant. Both variables u and p are best fit as constants with $u = 1.7534$ and $p = .8767$. The F-ratio of 9.23 for this fit, which accounts for the contribution made by the covariate minutes observed, is significat at $p < .03$. The contribution of

TABLE 6.3

ANALYSIS OF VARIANCE: H1.1a

Variable	Reg. Coeff.
u	-1.5878
p	1.0000 (subj. to $p \leq 1$)

Source of Variation	SS	df	MS	F-ratio	P
Total	614.00	12	--	--	--
u	533.34	1	--	--	--
p (given u)	69.78	1	69.78	41.41	< .005
$u_{ij} - u$ and $p_{ij} - p$	4.15	6	.69	0.41	--
Residuals	6.74	4	1.68	--	--

$u_{ij} - u$ and $p_{ij} - p$, the additive facotrs to accout for pretest, treatment, and interaction, is not significant. The analysis fails to reject part H1.1b. This is summarized in TABLE 6.4.

The same model is used to test hypothesis part H1.1c, the effect of pretest, treatment, and interaction on minutes satisfying criterion 2 with minutes observed taken as a covariate. In this case neither minutes observed nor pretest, treatment, and interaction contribute significantly to the sum of squares, the former adding .05 and the latter 5.70. This analysis fails to reject part H1.1c and also rejects minutes observed as a significant factor. A summary of this analysis is in TABLE 6.5.

TABLE 6.4

ANALYSIS OF VARIANCE: H1.1b

Variable	Reg. Coeff.
u	-1.7534
p	.8767

Source of Variation	SS	df	MS	F-ratio	P
Total	424.00	12	--	--	--
u	341.33	1	--	--	--
p (given u)	51.43	1	51.43	9.23	<.03
$u_{ij} - u$ and $p_{ij} - p$	9.00	6	1.50	.27	--
Residuals	22.24	4	5.56	--	--

Criterion 2 as written can be satisfied only by verbal behaviors. Background noise during the videotaping made it very difficult for the scorers to hear what the students said. The shyness or reluctance to verbalize shown by some students meant that they hardly talked at all. Such factors indicate that criterion 2 as presently written is not a good measure of mathematical creativity. Consequently only this one analysis is performed using the scores on criterion 2.

TABLE 6.5

ANALYSIS OF VARIANCE: H1.1c

Source of Variation	SS	df	MS	F-ratio	P
Total	21.00	12	--	--	--
u	14.08	1	--	--	--
p (given u)	.05	1	.05	.17	--
$u_{ij} - u$ and $p_{ij} - p$	5.70	6	.95	3.28	< .21
Residuals	1.17	4	.29	--	--

Hypothesis part H1.1d, the effect of pretest, treatment, and interaction on minutes satisfying criterion 5 with minutes observed taken as a covariate, is also tested using the linear model described earlier. Neither minutes observed, which adds 2.96 to the sum of squares, nor pretest, treatment and interaction, which add 7.22, make significant contributuions. This analysis fails to reject part H1.1d and also rejects minutes observed as a significant factor. A summary of this analysis is in TABLE 6.6.

The scoring of criterion 5 as the number of minutes in which the student achieved a mathematically elegant product is not as appropriate to the spirit of that criterion, in the author's opinion, as would be counting the number of products the student achieved during the total time he was observed. Unfortunately this realization came too late to change the way the scoring was done. For this reason, no further analyses are made using the scores on criterion 5.

TABLE 6.6

ANALYSIS OF VARIANCE: H1.1d

Source of Variation	SS	df	MS	F-ratio	P
Total	119.00	12	--	--	--
u	102.08	1	--	--	--
p (given u)	2.96	1	2.96	1.76	--
$u_{ij} - u$ and $p_{ij} - p$	7.22	6	1.20	.71	--
Residuals	6.74	4	1.68	--	--

The second experimental hypothesis, H1.2, concerns gain scores of those students who were pretested.

H1.2: There is not a positive effect of treatment on gain scores in the ratio of the number of minutes the student satisfies a criterion or the pseudocriterion to the number of minutes the student is observed.

Hypothesis H1.2 is divided into two parts: H1.2a concerning the pseudocriterion and H1.2b concerning criterion 3.

The model for testing the parts of sub-hypothesis H1.2 is an analysis of variance with the gain scores as the dependent variable and treatment as the independent variable. The gain scores were computed by first calculating for both pretest and posttest the ratio of the number of minutes satisfying the pseudocriterion or criterion 3 to the number of minutes observed and then subtracting the pretest ratio from the posttest ratio.

The analyses of variance performed on part H1.2a, the effect of treatment on gain scores in the pseudocriterion, shows that treatment, with a standard regression coefficient of 0.1485, does not contribute a large sum of squares. The F-ratio is 0.09 which is not significant. The analysis fails to reject sypothesis part H1.2a. This analysis is summarized in TABLE 6.7.

TABLE 6.7

ANALYSIS OF VARIANCE: H1.2a

Variable			Stand. Reg. Coeff.
Treatment			0.1485
Source of Variation	**SS**	**df**	**MS**
Total	0.26	5	--
Regression	0.01	1	0.01
Residuals	0.25	4	0.06
F-ratio: 0.09			

The analysis of variance made on part H1.2b, the effect of treatment on gain scores in criterion 3, shows that treatment, with a standard regression coefficient of 0.1115 does not contribute a large sum of squares. Since the F-ratio is 0.05, which is not significant, the analysis fails to reject hypothesis part H1.2b. This analysis is summarized in TABLE 6.8.

TABLE 6.8

ANALYSIS OF VARIANCE: H1.2b

Variable			Stand. Reg. Coeff.
Treatment			0.1115
Source of Variation	**SS**	**df**	**MS**
Total	0.51	5	--
Regression	0.01	1	0.01
Residuals	0.50	4	0.13
F-ratio: 0.05			

The third experimental hypothesis adds as a covariate to the second experimental hypothesis the pretest ratio of number of minutes satisfying to number of minutes observed.

H1.3: There is not a positive effect of treatment on gain scores in the ratio of the number of minutes the student satisfies a criterion or the pseudocriterion to the number of minutes the student is observed when the corresponding ratio for the pretest is taken as a covariate.

Hypothesis H1.3 has two parts: H1.3a concerning the pseudocriterion and H1.3b concerning criterion 3.

The model for testing the parts of hypothesis H1.3 is a modification of the one used for testing the parts of hypothesis H1.2.

The change is the addition of the corresponding pretest ratio as an independent variable.

The analysis of variance made on part H1.3a, the effect of treatment on the gain scores in the ratios pertaining to the pseudocriterion with the pretest ratio taken as a covariate, fails to reject this part of the hypothesis. The standard regression coefficients are -0.5224 for the pretest ratio and 0.0477 for treatment, but the F-ratio is 0.60 which is not significant. This analysis is summarized in TABLE 6.9.

TABLE 6.9

ANALYSIS OF VARIANCE: H1.3a

Variable	Stand. Reg. Coeff.
pretest ratio	-0.5224
treatment	0.0477

Source of Variation	SS	df	MS
Total	0.26	5	--
Regression	0.08	2	0.04
Residuals	0.19	3	0.06

F-ratio: 0.60

A similar situation results from the analysis of variance performed on part H1.3b, the effect of treatment on the gain scores in the ratios pertaining to criterion 3 with the pretest ratio taken as a covariate. The standard regression coefficients are -0.4932 for the pretest ratio and 0.1703 for treatment. The F-ratio of 0.51 is not significant, so the analysis fails to reject that hypothesis part. This analysis is summarized in TABLE 6.10.

TABLE 6.10

ANALYSIS OF VARIANCE: H1.3b

Variable	Stand. Reg. Coeff.
Pretest ratio	-0.4932
Treatment	0.1703

Source of Variation	SS	df	MS
Total	0.51	5	--
Regression	0.13	2	0.06
Residuals	0.38	3	0.13

F-ratio: 0.51

It had been originally planned to test an experimental hypothesis concerning the effect of pretest, treatment, and interaction on the number of minutes a student satisfied criterion 3 or the pseudocriterion when the general creativity scores of the student and the number of minutes the student was observed are taken as covariates. The model to test this hypothesis is an extension of the one used to test hypothesis H1.1. Unfortunately, this model has so many independent variables that achieving a meaningful fit with only twelve subjects is not possible.

None of the experimental hypotheses were rejected. This means that pretest, treatment and interaction did not produce statistically significant positive changes in the mathematical creativity scores. The only factor which did contribute significantly to the number of minutes satisfying the pseudocriterion or criterion 3 is the number of minutes observed. This completes the analyses made on the mathematical creativity scores; the analyses using the general creativity scores are presented next.

Analyses Performed on the General Creativity Scores

Nine analyses were performed to test various aspects of general hypothesis H2.

H2: Participation in the program will not affect a student's performance on a test of general creative ability.

This general hypothesis is devided into three experimental hypotheses, each of which is further divided into three parts corresponding to the dimensions of fluency, flexibility, and originality. As with the aspects of mathematical creativity, the method used to test the parts of the sub-hypotheses concerned with general creativity is analysis of variance.

The first experimental hypothesis postulates the effects of pretest, treatment, and their interaction.

H2.1: There is no significant effect of pretest, treatment, or their interaction on the posttest general creativity scores.

This hypothesis is divided into three parts: H2.1a concerning fluency scores, H2.1b concerning flexibility scores, and H2.1c concerning originality scores.

The analysis for hypothesis part H2.1a, the effect of pretest, treatment and their interaction on fluency posttest scores, gives standardized regression coefficients of -0.1509 for pretest, -0.4864 for treatment and 0.5535 for interaction. The F-ratio for the regression is 3.47 which is marginally significant, $p < .08$. The group averages to the nearest whole number (from TABLE 6.2) are 55 for both pretested groups, 51 for the group that received treatment but no pretest, and 62 for the group that received neither pretest nor treatment. The high average for the latter group accounts for the results of the regression. Since there were only three subjects in each experimental group it is possible that this difference is due primarily to the small sample size. This analysis is summarized in TABLE 6.11.

For hypothesis part H2.1b, the effect of pretest, treatment, and interaction on posttest flexibility scores, the standardized regression

TABLE 6.11

ANALYSIS OF VARIANCE: H2.1a

Variable				Stand. Reg. Coeff.
Pretest				-0.1509
Treatment				-0.4864
Interaction				0.5535
Source of Variation	**SS**	**df**	**MS**	
Total	296.25	11	--	
Regression	167.58	3	55.86	
Residuals	128.67	8	16.08	

F-ratio: 3.47; $p < .08$

coefficients are 0.4961 for pretest, -0.2417 for treatment, and 0.4198 for interaction. The F-ratio for the regression is 2.47 which is not significant, $p < .16$, so the analysis fails to reject part H2.1b. This analysis is summarized in TABLE 6.12.

The analysis made on hypothesis part H2.1c, the effect of pretest, treatment, or interaction on posttest originality scores, fails to reject that hypothesis part at a significant level. The standardized regression coefficients are 0.2823 for pretest, -0.4436 for treatment and 0.3952 for interaction with an F-ratio for the regression of 2.03 which is not significant, $p < .21$. This analysis is summarized in TABLE 6.13.

TABLE 6.12

ANALYSIS OF VARIANCE: H2.1b

Variable				Stand. Reg. Coeff.
Pretest				0.4961
Treatment				-0.2417
Interaction				0.4198
Source of Variation	**SS**	**df**	**MS**	
Total	514.92	11	--	
Regression	247.57	3	82.53	
Residuals	267.33	8	33.42	

F-ratio: 2.47; $p < .16$

The second experimental hypothesis concerning the general creativity scores involves analysis of the gain scores of those students who were pretested.

H2.2: There is no significant effect of treatment on general creativity gain scores.

This hypothesis is divided into three parts: H2.2a concerning fluency gain scores, H2.2b concerning flexibility gain scores, and H2.2c concerning originality gain scores.

TABLE 6.13

ANALYSIS OF VARIANCE: H2.1c

Variable			Stand. Reg. Coeff.
Pretest			0.2823
Treatment			-0.4436
Interaction			0.3952

Source of Variation	SS	df	MS
Total	1280.92	11	--
Regression	554.25	3	184.75
Residuals	726.67	8	90.83

F-ratio: 2.03 ; $p < .21$

The parts of this hypothesis are tested by analysis of variance using treatment as the independent variable and the gain scores as the dependent variable. Gain scores are computed by subtracting pretest scores from posttest scores.

The analysis performed on part H2.2a, the effect of treatment on fluency gain scores, gives a standardized regression coefficient of -0.1689 for the treatment. The F-ratio for the regression is 0.12 which is not significant, so the analysis fails to reject part H2.2a at a significant level. This analysis is summarized in TABLE 6.14.

TABLE 6.14

ANALYSIS OF VARIANCE: H2.2a

Variable				Stand. Reg. Coeff.
Treatment				-0.1689
Source of Variation	SS	df	MS	
Total	374.06	5	--	
Regression	10.67	1	10.67	
Residuals	363.33	4	90.83	
F-ratio: 0.12				

Hypothesis part H2.2b, the effect of treatment on flexibility gain scores, is not rejected by the analysis, because the F-ratio of 0.04 for the regression is not significant. The standardized regression coefficient for the treatment is 0.0931. This analysis is summarized in TABLE 6.15.

The analysis made on hypothesis H2.2c, the effect of treatment on originality gain scores, fails to reject that part. The standardized regression coefficient for the treatment is -0.4355; the F-ratio of 0.94 for the regression is not significant. This analysis is summarized in TABLE 6.16.

The third experimental hypothesis on general creativity scores adds the pretest scores as a covariate to the model tested in the second experimental hypothesis.

TABLE 6.13

ANALYSIS OF VARIANCE: H2.2b

Variable				Stand. Reg. Coeff.
Treatment				0.0931
Source of Variation	**SS**	**df**	**MS**	
Totals	309.33	5	--	
Regression	2.67	1	2.67	
Residuals	303.67	4	76.17	

F-ratio: 0.04 .

H2.3: There is no significant effect of treatment on general creativity gain scores with pretest scores taken as a covariate.

As with the previous experimental hypotheses, this one is divided into three parts: H2.3a concerning fluency gain scores, H2.3b concerning flexibility gain scores, and H2.3c concerning originality gain scores.

The analysis of variance model used to test the parts of hypothesis H2.3 has the gain scores as the dependent variable and pretest scores and treatment as the independent variable.

The analysis made on part H2.3a, the effect of treatment on fluency gain scores with fluency pretest scores taken as a covariate,

TABLE 6.16

ANALYSIS OF VARIANCE: H2.2c

Variable				Stand. Reg. Coeff.
Treatment				-0.4355
Source of Variance	SS	df	MS	
Total	425.33	5	--	
Regression	80.67	1	80.67	
Residuals	344.67	4	86.17	

F-ratio: 0.94

gives standardized regression coefficients of -0.8334 for the fluency pretest scores and 0.0480 for treatment. The F-ratio of 3.13 for the regression is not significant, $p < .23$, so the analysis fails to reject part H2.3a. This analysis is summarized in TABLE 6.17.

For part H2.3b, the effect of treatment on flexibility gain scores with flexibility pretest scores taken as a covariate, the standardized regression coeffieients are -0.7683 for flexibility pretest scores and 0.1742 for treatment. The F-ratio for the regression is 2.18 which is not significant, $p < .32$, so the analysis fails to reject hypothesis part H2.3b. This analysis is summarized in TABLE 6.18.

TABLE 6.17

ANALYSIS OF VARIANCE: H2.3a

Variable				Stand. Reg. Coeff.
Fluency Pretest Scores				-0.8334
Treatment				0.0480
Source of Variation	**SS**	**df**	**MS**	
Total	374.00	5	--	
Regression	252.84	2	126.42	
Residuals	121.16	3	40.39	
F-ratio: 3.13 ; p < .23				

The analysis made on hypothesis part H2.3c, the effect of treatment on originality gain scores with originality pretest scores taken as a covariate, fails to reject that part. The standardized regression coefficients are -0.7536 for originality pretest scores and 0.0394 for treatment. The F-ratio is 1.71 which is not significant. This analysis is summarized in TABLE 6.19.

Only one of the nine analyses tends to reject the general hypothesis that participation in the program did not affect general creativity scores. That analysis, showing negative interaction effects of pretest and treatment on fluency scores, is marginally significant, $p < .08$, and seems due to a high average fluency score

TABLE 6.18

ANALYSIS OF VARIANCE: H2.3b

Variable	Stand. Reg. Coeff.
Flexibility Pretest Scores	-0.7683
Treatment	0.1742

Source of Variation	SS	df	MS
Total	307.33	5	--
Regression	182.08	2	91.04
Residuals	125.25	3	41.75

F-ratio: 2.18; $p < .32$

for the group which received neither pretest nor treatment. This result may be due to the small number of subjects involved in the experiment.

In addition to examining the data to see if they do or do not support various aspects of the two general hypotheses, a correlation matrix was calculated for the two primary mathematical creativity scores and three general creativity posttest scores used to test the hypotheses. Because only twelve subjects were involved in the post-tests, most of the correlations between the scores are not significantly different from zero, $p > .1$. The two which are significantly different from zero, both at level $p < .002$, are .824 for the

TABLE 6.19

ANALYSIS OF VARIANCE: H2.3c

Variable	Stand. Reg. Coeff.
Originality Pretest Scores	-0.7536
Treatment	0.0394

Source of Variation	SS	df	MS
Total	425.33	5	--
Regression	226.31	2	113.15
Residuals	199.03	3	66.34

F-ratio: 1.71 ; $p < .45$

correlation between the flexibility and originality scores and .923 for the correlation between satisfying the pseudocriterion and satisfying criterion 3. Torrance reports intercorrelations above .70 for fluency, flexibility and originality scores for Figural Form A, which was used as the posttest. These intercorrelations were made from test scores of 48 Wisconsin second grade students (Torrance 1966f, p. 82). It is likely that the lower number of test scores could account for the lower correlations in the two cases. The high correlation between scores on the pseudocriterion and criterion 3 is to be expected since the pseudocriterion is considered satisfied if at least one criterion is satisfied and criterion 3 is the criterion most frequently considered satisfied. Thus this correlation reflects a part--whole relationship. It is interesting to note that all correlations between mathematical and general creativity scores are negative. The correlation matrix is given as TABLE 6.20.

TABLE 6.20

CORRELATION MATRIX OF POSTTEST SCORES

	flex	orig	any	rel
flu	.289	.360	-.052	-.204
flex		.824	-.190	-.137
orig			-.269	-.171
any				.923

It remains to interpret the results of the statistical analyses in terms of the present experiment to indicate directions for future investigations. This is done in the next chapter.

Chapter VII

CONCLUSIONS

7.1 OUTLINE OF CHAPTER VII

In this chapter conclusions are drawn about the test instrument and the results of the experiment. The test instrument is discussed first, in terms of its strengths and weaknesses, and some proposals are made for its improvement. Then the experiment is examined, the results of the statistical analyses are interpreted, alternative hypotheses are offered, and implications for future research are indicated. A summary of the thesis concludes the chapter.

7.2 THE TEST INSTRUMENT

A test instrument was developed to measure observable mathematical creativity. One part of the instrumant is fixed--the criteria which describe aspects of mathematical creativity in terms of observable behaviors. The other part of the test instrument in the mathematical problem on which the person being tested works. This problem can be chosen to fit the needs of an experiment . In the experiment reported in this thesis, trained scorers used the criteria to judge the activity of the person as he worked.

The criteria were developed by the author and passed a face validation procedure in which they were judged by seven Professors of Mathematics at the University of Wisconsin. The criteria were written in an expanded form for the validation and were subsequently compactified for easier use by the scorers. The rewritten criteria are listed here; the relevant glossary is in Appendix D. It is the belief of the author that the rewriting was consonant with the criteria as approved by the Professors of Mathematics and with the comments made by them.

1. Introducing a goal: In the absence of a specific stated mathematical goal, the student either verbally suggests or exhibits goal-directed behavior with respect to some appropriate goal.

2. Identifying a property: The student states an appropriate unstated property of the task.

3. Seeking a relationship: The student conjectures, states, demonstrates, or investigates a possible relationship between (a) some appropriate property of the task he is pursuing and (b) either some other appropriate property of the same task or some appropriate property of some different task.

4. Seeking a generalization: The student conjectures, states, demonstrates, or attempts a possible appropriate generalization.

5. Reaching a mathematically elegant product: The student achieves, states, or demonstrates a mathematically elegant and appropriate product.

6. Modifying the task: After having pursued the task as outlined by the teacher, the student either verbally suggests or pursues an appropriate unstated modification of the task.

The fifth criterion concerns the nature of the result of an activity; the other five criteria describe actions which may or may not lead to a worthwhile result, but which in themselves are aspects of mathematically creative activity because they describe activities which are a normal part of the preparation and manipulation stages in a mathematician's work. The first criterion, setting a goal for oneself, is partly a recognition of the motivational forces involved in the creative process as well as a description of some preparation and manipulation activities. The second, third, and fourth criteria involve observable aspects of the process by which one notices a mathematical property of something, seeks a relationship between the values of two mathematical properites in a specific case, and seeks a general setting in which a value of some property or a particular relationship exists. The sixth criterion outlines a process by which one might try to use the new idea in a familiar context or explore the further possibilities of the new idea.

The problems used in the experiment satisfied several requirements some of which can be stated in general terms. The problems must be suited to the mathematical sophistication of the person being tested. They must tend to generate observable behaviors during the process of solution. If the medium of videotape is used for observation, as was done in the experiment, then the problems should not tend to generate either actions of a large and small scope in a rapidly alternating sequence or solutions which are not planar in their essential details.

The experiences of the experiment comment on the scoring procedure and indicate ways in which the test instrument could be improved.

Interscorer agreement during use of the test instrument was reasonably high. This conclusion is based in part on the high overall average interscorer agreement (.87) and the reasonable to high overall interscorer agreements on five of the six criteria and on the pseudocriterion (range of .61 to 1.00). Only with respect to criterion 3 is the overall interscorer agreement low (.43). Actions on the videotapes viewed by the scorers made the scoring of criterion 3 difficult because, in the judgment of the author, many actions seemed either to almost satisfy criterion 3 or to just fail to satisfy that criterion. If under these circumstances, interscorer agreement drops to .26 as it did for criterion 3 in the scoring of the posttest tapes, then the test instrument would not be very valuable. However, it seems that much of the disagreement can be attributed to the scoring of criterion 3 done by one scorer; he said afterwards that he was consciously not following the guidelines set for scoring. This indicated that if the training techniques are improved and if the scorers follow the guidelines given, then reasonable to high interscorer agreements could be expected. In examining the interscorer reliabilities of his Tests of Creative Thinking Ability, Torrance found a similar situation: Low interscorer reliabilities occurred only when some scorers failed to follow the guidelines in the scoring manuals (Torrance 1966f, p. 19).

The test instrument is capable of measuring differences in the amount of observable mathematical creativity demonstrated on the videotapes, and therefore is appropriate for use in experimental situations in which comparison of scores is necessary for the statistical analyses. However, the experiences of the experiment indicate that some changes in the test instrument would improve its usefulness.

One indicated modification of the test instrument involves the blocking of the tapes into equal intervals. In the experiment, the tapes were blocked into one minute intervals starting with the beginning of the taped interaction of the student, the author, and the problem, and continuing until the end of that interaction. This method was not without problems in that the beginning and end minutes of the taped observation usually involved such activities as the explanation of the task by the author and the decision as to whether the student wished to work longer on the problem or was satisfied with his results and wished to return to his classroom. These activities varied in length from student to student and are not necessarily activities on which the student should be observed for the purpose of scoring his mathematical creativity. Some consistent method should be developed of elimination of "dead time" from the tapes used for scoring before blocking the tapes into equal time intervals.

A change in the length of the equal time intervals should be explored. If these intervals were of a shorter duration than one minute, for example twenty or thirty seconds, the test instrument might be able to make significantly greater discriminations among the persons observed. Although this kind of effect would be partially attributable to the fact that the number of observations made per total time would be higher using a shorter interval, it could also result from different scores being given to the person who exhibits an action briefly at approximately one minute intervals and the person who exhibits that action continuously.

Because the length of observation time varied among students and between pretest and posttest, the statistical analyses using the scores of the test instrument had to take this variation into consideration. If some arbitrary time limit were set so that each person being tested would be observed for the same length of time, then the statistical analyses would become simpler. This might have the effect of making the instrument more suitable for use in controlled experiments; the only significant factor contributing to the number of minutes satisfying a criterion or the pseudocriterion in the present experiment is the number of minutes observed. It does seem contrary to the nature of creativity, however, to set time limits. This matter should be further investigated to see if the apparent conflict between the nature of creativity and the desire for a simpler set of scores can be resolved.

Criterion 2 as now written can be satisfied only by verbal behavior. Unless some effort is made to assure that all subjects are equally likely to express themselves verbally, comparison of scores on criterion 2 will probably reflect differences in the tendency to verbalize as well as differences in one aspect of mathematical creativity. This indicates that a rewriting of criterion 2 or a restructuring of the criteria to absorb criterion 2 into the other criteria is necessary if the scores from the test instrument are to be used as direct measures of mathematical creativity.

Another indicated change involves criterion 5, the achievement of a mathematically elegant product. The other five criteria describe actions; criterion 5 describes a result. It is reasonable to score the other five criteria by dividing the time the student is observed into equal intervals, recording whether or not at least one of the student's actions satisfy each criterion during each interval, and disregarding any multiplicity of satisfying actions in the scoring of the interval. However, the achievement of a product is not an action but a result; it is more in the spirit of criterion 5 to give the student credit for each and every mathematically elegant product he achieves, regardless of how those products occur with respect to the time intervals. If one student achieves one product during an interval and another student achieves three products, it is only reasonable and fair to award the two students different scores. This modification in the scoring procedure should not be difficult to make; it was used by several of the scorers during the scoring of each minute on the experimental videotapes.

The choice of the equal interval method of scoring with disregard of any multiplicity of satisfying actions was made on the basis of one particular advantage of that method. The test instrument, in order to be valuable, must be usable by more than one scorer. One indication that it is usable by a wide range of trained scorers would be high interscorer agreement on what is observed. It was assumed that the equal interval method would insure that the scorers were observing and evaluating the same actions. The existence of a multiplicity of satisfying actions in many of the minutes taped means that in fact, each scorer could have been considering a different action when he marked a criterion as satisfied. Thus the method may produce seeming agreement when there is disagreement.

Another problem with this method is raised by the discussion of its appropriateness for scoring criterion 5. Perhaps satisfying actions would be more in the spirit of the other five criteria also. In fact, the method used does count a multiplicity of one sort--the number of minutes during which the satisfying action or actions occur. But is the meaning of a criterion like "seeking a relationship" based in the time spent in this activity or the number of different relationships sought or both? The author feels that both aspects are important but that the method used thus far really only measures one of them. A more complicated scoring procedure in which, for example, the number of different relationships sought and the time spent in seeking them were both recorded should be attempted. One would anticipate that interscorer agreement on a more complicated procedure might be lower than on the present one, but if this occurred, it would be compensated by the fact that the scores would be more meaningful.

The individual scores from the pretest and posttest (presented in TABLE 6.2) suggest an interesting conjecture. On those aspects of mathematical creativity which were scored non-zero in both testing sessions, the scores of the posttest average much higher than the scores of the pretest, regardless of the experimental group, even though the posttest was, on the average, a shorter task. Some aspects were scored non-zero in one task only. Perhaps the individual scores are primarily comments on the different natures of the two problems used. The fact that neither pretest nor treatment seemed to have influenced the scores in a statistically significant way adds strength to this conjecture. This suggests that before using the test instrument in another experiment, some evaluation of the properties of various problems should be made.

In summary, it can be said that the test instrument appears to be basically sound. The criteria do describe aspects of observable mathematical creativity and can be used reliably in one scoring method to evaluate observed behaviors. Some further exploration is needed to determine the best way to use the criteria to score actions.

7.3 THE EXPERIMENT

An experiment was conducted in which the test instrument just discussed was used to measure the effects of a treatment in the form of an instructional program consisting of a sequence of open-ended mathematical problems developed to encourage the individual mathematical creativity of first grade students. The Torrance Tests of Creative Thinking, Figural Forms A and B, were used to measure the effects of the treatment on general creativity.

Two general hypotheses were tested in the experiment.

H1: Participation in the experimental program will increase a student's observable mathematical creativity.

H2: Participation in the experimental program will not affect a student's performance on a test of general creativity.

Both of these general hypotheses were divided into several experimental hypotheses, each of which was tested using each of the appropriate scores. The statistical analyses are presented in Chapter VI; in this chapter the results of those analyses are interpreted.

Interpretation of Results on Mathematical Creativity

The statistical analyses testing aspects of general hypothesis H1 tend toward rejection of that hypothesis in favor of the null hypothesis: Participation in the program had no effect on mathematical creativity scores.

Of the eight analyses pertaining to general hypothesis H1, four were calculated using gain scores in the ratio of the number of minutes the student satisfied the pseudocriterion or criterion 3 to the number of minutes the student was observed. This ratio was used as a means of comparing pretest and posttest scores because the number of minutes observed varied between pretest and posttest and among students. None of these four analyses found significant contributions to the variance in the gain scores made by either treatment alone or treatment with pretest scores as a covariate. It is possible that the small variance in the gain score ratios contributed to the lack of statistically significant results from these analyses.

The other four analyses pertaining to hypothesis H1 reject experimental hypothesis H1.1, that the pretest, treatment, and interaction had a positive effect on the mathematical creativity scores. In two of the analyses, those concerning the pseudocriterion and criterion 3, significant influence on the number of minutes satisfying was made by the number of minutes the student was observed.

These analyses indicate rejection of hypothesis H1 and thus failure to reject the corresponding null hypothesis. There are some possible factors unaccounted for in this indicated rejection. It is possible that weakness in the test instrument prevented or obliterated measurement of effects of the pretest, treatment, and interaction. It may be inherent in the nature of creativity that the behavior sought is not stable over time and that more observations, adequately spaced, be made. Difficulties with the videotape equipment at the posttesting sessions also may have contributed. Three students were observed for less time than they actually worked due to taping problems. Of the three shortened observations, two are missing at most two minutes of the actual time and one lacks several minutes, over half of the time worked; all the missing times are from the ends of the sessions. Unfortunately, two of the shortened observations--the major one and one of the minor ones--occur in the group of three students which received both pretest and treatment. Since in most cases the student produced more as he became engrossed in the problem, the loss of the later minutes of a testing session may vary well prejudice the scores to be lower than they would have been if the entire session were taped. This effect might persist even if ratios rather than raw scores were used. Another factor which could contribute to the lack of significant differences in the small number of students involved in the experiment, since the larger the number of subjects, the more significant small changes become, and the possibility that the program did produce small changes has not been ruled out.

The rejection of hypothesis H1.1 is so overwhelming in one case in favor of a significant contribution made by the number of minutes observed ($p<.005$), that one is inclined to suspect that a statistically significant rejection might have occurred even without technical and measurement problems. The statistical rejection of hypothesis H1 is supported by the casual observation made by the author during the posttesting sessions. This leads to an examination of the experimental program and its assumptions in a search for alternative hypotheses, each of which deserves investigation.

In order to facilitate cooperation with the schools in testing an experimental program, and because the program was a first step, several decisions concerning timing of the lessons and length of the program were made and not experimentally investigated. Perhaps a sequence of fifteen daily lessons of twenty minutes duration is not the best timing of an attempt to encourage mathematical creativity. Three factors should be considered: the duration of the individual lessons, the frequency of the lessons, and the total time over which the program is used. One alternative is twice weekly lessons of an hour each for two months. A longer lessons would mean that less time is taken up by review and clean-up activities and would allow for greater development of the mathematical aspects of the problem situations. In order to compensate for the longer absence of the students from the regular class, daily lessons might have to be abandoned in favor of less frequent ones. Since it is possible that the experimental program did not produce significant changes because the total time over which it was used was too short, an increase in total time might be helpful.

It is possible that not enough emphasis was placed on imparting general principles to the students. There was a hesitancy to rely on verbal communication because it was felt that understanding abstract verbal principles and verbalizing ideas were difficult activities for first grade students and might interfere with the mathematical investigations taking place. The "simple" episode is an example of some success in communicating a verbal principle to the students and indicates that possibly more of this kind of activity could be done than was originally thought. The advantage of communicating general principles is, of course, based on the assumption that these principles will readily transfer to new situations, such as a posttest.

Another possible reason for the lack of statistically significant positive effects due to the experimental progran can be inferred from some of the events during the lessons and pilot studies. Although it was not a hypothesis of the experiment, one question which can be answered on the basis of the experiment is whether first grade students can exhibit any behavior which contains aspects of mathematically creative activity. The account of the program and the events of the pilot studies clearly indicate that the answer to this question must be yes. The students in the program set goals for themselves and followed them through, for example making an octahedron; they stated (in their own words) such properties as having rows or being open. The two methods used to try to duplicate the octahedron clearly show investigation of relationships. The student who taped the card stock shapes together to make a covering and then pointed out to the teacher that this activity was similar to the previous tiling one had made a generalization. The chart showing the new shapes which could tesselate as discovered by the students is an example of mathematically elegant products achieved by them. An interesting example of a modification occurred in the first pilot study, as the second response to the problem of six straws. In short, ample evidence exists that under the kind of conditions described in Chapter II first grade students can exhibit behavior satisfying any and all of the six criteria which describe aspects of mathematically creative activity.

A program might be prevented from having a significant effect because mathematically creative behaviors can be elicited from first grade students rather readily. It may be that these behaviors are natural and frequent among first grade students, given the proper conditions. If this is the case, then any attempt to increase the likelihood of these behaviors could be likened to an attempt to increase the scores of persons who consistently score high--such an increase may be possible, but it would come only after expenditure of mammoth amounts of time and energy. This magnitude of effort was clearly not built into the program.

It is possible that the lack of significant effect of the program was due to defects in the program itself, and not to either problems with the test instrument or the degree of mathematical creativity exhibited by first grade students. Although harboring no illusions about the experimental program's level of perfection, the author wonders whether efforts directed toward first grade students may be necessarily doomed. If this were to be found true, after the experimentations with improved programs for first grade students, then one would need to turn to some new directions for future research. Such a direction might be to examine the behavior of older students.

A common experience among college teachers is that mathematical creativity is not easily elicited from their students; somewhere between first grade and freshman year the degree of mathematical creativity exhibited by the students should be such that a moderate program could be expected to produce significant results. Once a higher grade level has been found at which a program has a significant effect, two choices are open. One could test various programs at that level to see which produced the most effect and attempt to perfect such a program. Alternatively, one could return to the first grade students and run a long term experiment to determine the effects of regular exposure to situations which encourage mathematical creativity starting in the first grade and continuing until the higher grade level. The author feels that the latter choice would be more valuable because it could more easily be adapted to curriculum reform and is less remedial in nature.

The hypothesis of primary interest was the one concerned with mathematical creativity. The results of the analyses testing the other hypothesis, concerning general creativity are also of interest.

Interpretation of Results on General Creativity

Only one of the analyses testing hypothesis H2 showed any significant effect of treatment, i.e. participation in the program, on general creativity scores. Although that analysis showed a negative interaction effect of pretest and treatment on fluency scores, it is only marginally significant, $p<.08$, and may be due to the small number of subjects involved in the experiment. Because this result is only marginal and because the other analyses, including others on fluency scores, do not show any significant effects of treatment, it can be said that these analyses therefore fail to reject hypothesis H2 which is the null hypothesis: Participation in the program will not affect general creativity scores. This is the expected result since no effort was made during the program to exercise or otherwise directly improve the general creative abilities.

The correlations between the general creativity scores and the mathematical creativity scores, while not significantly different from zero, are interesting in that they are all negative. This indicated direction of correlation could be tested experimentally in a status study using a large number of persons in order that small negative correlations would have statistical significance. The exact interpretation of such a correlation, if it were found to hold, is unclear, but it would be an interesting result.

7.3 SUMMARY

The importance of mathematical creativity is widely acknowledged. In this thesis, some characteristics of the creative process and the creative person were examined. On the basis of this background, six criteria describing observable aspects of mathematical creativity were identified. These criteria were face validated be seven Professors of Mathematics at the University of Wisconsin and serve as part of a test instrument to measure observable mathematical creativity. One set of conditions conducive to mathematical creativity was proposed and activities which satisfy these conditions were piloted. From these activities were developed both an instructional program to encourage individual mathematical creativity in first grade students and problems to use as part of the test instrument. An experiment was conducted to determine the effects of participation in the program on observable mathematical creativity, these effects were measured using the test instrument developed in this thesis. The effects on general creativity were measured using the Torrance Tests of Creative Thinking, Figural Forms A and B.

The literature suggests that evidence of mathematical creativity could be obtained by observing the actions of a person while he works on a mathematics problem in addition to the more traditional means of evaluating the results he achieves. In this thesis, a test instrument was developed which measures aspects of the observable mathematical creativity demonstrated by a person working on a mathematics problem through use of six criteria against which the person's behavior is evaluated. Five of the six criteria describe activities in which a person might be engaged as he pursues a problem; the sixth criterion describes the result of those activities.

The literature also suggests ways in which to encourage mathematically creative activity. One kind of mathematics situation which seems suitable to this encouragement is an open-ended problem having a moderate amount of structure, suited to the mathematical sophistication of the person, and generating observable behaviors during the process of solution. This last requirement was placed on the problem because it helps a teacher follow the process of solution without interfering with it. The literature suggests that the presence of an interested, creative teacter who is sensitive to the goals of the student and has an accepting attitude serves to encourage creative efforts.

Two pilot studies were conducted to help interpret the above requirements and to develop activities suitable for first grade students. From the successful activities a program of fifteen daily lessons, each of twenty minutes duration, was constructed. For the purposes of the program, all activities were required to have similar content: the incidence-type relationships pertaining to arrangements of triangles. Two of the successful activities were used as part of the test instrument.

An experiment was conducted to measure the effects on observable mathematical creativity and on general creativity of participation in this program to encourage mathematical creativity of firat grade students. The working hypotheses of the experiment were that participation in the program first, would increase a student's observable mathematical creativity and second, would not affect a student's general creativity. No significant effect of treatment was found on either kind of creativity, so the first hypothesis was rejected and the second hypothesis was not rejected.

There are three kinds of reasons which might explain the lack of statistically significant increases in observable mathematical creativity due to the program. At one level, measurement problems and the small number of subjects may have acted to mask any increases which did occur. Another possiple reason for the results could be that the aspects of timing of the program the the hesitancy to verbally impart general principles resulted in no change in behavior on the part of the students in the program. This alternative hypothesis could be tested by making the indicated improvements in the program and conducting another experiment.

A third possibility is indicated by the fact that the activities of the program readily elicited mathematically creative activity from the student participants. The author suspects that first grade students may exhibit such a high degree of observable mathematical creativity under the suitable conditions described in Chapter II that no moderate program might be able to increase the level of observable mathematical creativity of these students. If this were the case, then the next step in a research program might be a search for the grade level at which a program could produce a significant difference. Once this were determined, efforts could be directed at either perfecting a program for the higher grade level or testing the effects on students at the higher grade level of a long range program starting in the first grade. The latter type of effort could have direct consequences for curriculum reform.

The test instrument appears to be basically sound, The criteria do describe aspects of observable mathematical creativity and can be used reliable in one scoring method to evaluate observed behavior. Some further exploration is needed to determine the best was to use the criteris to score actions. Any improvements which may be made in the test instrument should be tested independently of the testing of a new program to encourage mathematical creativity in order to avoid in the future the problem of whether the lack of statistically significant differences between experimental groups is primarily due to differences not existing or not being measured.

The pilot studies and the experimental program give substantial evidence that under one set of suitable conditions first grade students can exhibit behavior satisfying all the criteria which describe aspects of mathematically creative activity. The major contributions of this thesis are the identification and validation of criteria describing observable aspects of mathematical creativity and the presentation of evidence that young students in first grade can exhibit behaviors satisfying these criteria.

REFERENCES

American Association for the Advancement of Science. 1965. An evaluation model and its application. Science--a process approach. AAAS Miscellaneous Publication 65-9.

Association of Teachers of Mathematics. 1968. Notes on mathematics in primary schools. Cambridge, England: University Press.

Cambridge Conference on School Mathematics. 1963. Goals for school mathematics. Boston: Houghton Mifflin.

Campbell, D. T., and Stanley, J. C. 1966. Experimental and quasi-experimental designs for research. Chicago: Rand McNally.

Crutchfield, R. A. 1967. Conformity and creative thinking. In Contemporary approaches to creative thinking, ed. H. E. Gruber, G. Terrell, and M. Wertheimer, pp. 120-140. New York: Atherton Press.

Crutchfield, R. A. and Covington, M. V. 1963. Facilitation of creative thinking and problem solving in school children. Paper read at Symposium on Learning Research Pertinent to Educational Improvement, American Association for the Advancement of Science, 29 December 1963 at Cleveland.

Davis, G. A. 1969. Thinking creatively in adolescence: a discussion of strategy. In Studies in adolescence, ed. R. E. Grinder, pp. 538-545. New York: Macmillan.

Flavell, J. H. 1963. The developmental psychology of Jean Piaget. New York: Van Nostrand.

Guha, S. R. 1966. Multiple regression analysis program. REGAN1. Madison: University of Wisconsin Computing Center.

Guilford, J. P. 1962. Creativity: its measurement and development. In A source book for creative thinking, ed. S. J. Parnes and H. F. Harding, pp. 151-168. New York: Scribners.

Hadamard, J. 1954. An essay on the psychology of invention in the mathematical field. New York: Dover.

Hendrix, G. 1961. Learning by discovery. The Mathematics Teacher, vol. 54, no. 5, pp. 290-299.

Howard, I. P. and Templeton, W. B. 1966. Human spatial orientation. New York: Wiley.

Kaiser Aluminum and Chemical Corporation. 1968. You and creativity. Kaiser Aluminum NEWS, vol. 25, no. 3. Oakland.

Klausmeier, H. J. and Goodwin, W. 1966. Learning and human abilities. New York: Harper and Row.

Myers, R. E. and Torrance, E. P. 1965a. Can you imagine? Boston: Ginn.

Myers, R. E. and Torrance, E. P. 1965b. Can you imagine? Teacher's guide. Boston: Ginn.

Myers, R. E. and Torrance, E. P. 1966a. For those who wonder. Boston: Ginn.

Myers, R. E. and Torrance, E. P. 1966b. For those who wonder. Teacher's guide. Boston: Ginn.

Ontario Institute for Studies in Education. 1967. Geometry. Kindergarten to grade thirteen. Toronto: Ontario Institute for Studies in Education.

Osborn, A. F. 1963. Applied imagination. New York: Scribners.

Piaget, J. 1964. Development and learning. In Piaget revisited, ed. R. E. Ripple and V. N. Rockcastle, pp. 7-20. Ithaca: Cornell University Press.

Poincaré, H. 1913. The foundations of science. Trans. by G. B. Halsted. New York: Science Press.

Polya, G. 1957. How to solve it. Garden City, N. Y.: Doubleday Anchor.

Polya, G. 1965. Mathematical discovery, volume II. New York: Wiley.

Smith, J. A. 1966. Setting conditions for creative teaching in the elementary school. Boston: Allyn and Bacon.

Torrance, E. P. 1962. Guiding creative talent. Englewood Cliffs, N. J.: Prentice Hall.

Torrance, E. P. 1966a. Thinking creatively with pictures. Booklet A. Princeton: Personnel Press.

Torrance, E. P. 1966b. Thinking creatively with pictures. Booklet B. Princeton: Personnel Press.

Torrance, E. P. 1966c. Thinking creatively with words. Booklet A. Princeton: Personnel Press.

Torrance, E. P. 1966d. Torrance tests of creative thinking. Directions manual and scoring guide. Figural test. Booklet A. Research edition. Princeton: Personnel Press.

Torrance, E. P. 1966e. Torrance tests of creative thinking. Directions manual and scoring guide. Verbal test. Booklet B. Research edition. Princeton: Personnel Press.

Torrance, E. P. 1966f. Torrance tests of creative thinking. Norms technical manual. Research edition. Princeton: Personnel Press.

Torrance, E. P. 1967. The Minnesota studies of creative behavior: national and international extensions. Journal of Creative Behavior 1: 137-154.

Torrance, E. P. 1968. Torrance tests of creative thinking. Directions manual and scoring guide. Figural test. Booklet B. Research edition. Princeton: Personnel Press.

Vygotsky, L. S. 1962. Thought and language. Ed. and trans. by E. Hanfmann and G. Vakar. Cambridge: M.I.T. Press.

Westcott, A. M. and Smith, J. A. 1967. Creative thinking of mathematics in the elementary school. Boston: Allyn and Bacon.